MODELO DE CLASIFICACIÓN DE ÍMAGENES BASADO EN MULTIVIEW LEARNING

Monografía

Kevin Camilo Díaz Monje

Universidad Distrital Francisco José de Caldas
Maestría en Ciencias de la Información y las Comunicaciones
Facultad de Ingeniería
Bogotá D. C., Colombia
2025

MODELO DE CLASIFICACIÓN DE IMÁGENES BASADO EN MULTIVIEW LEARNING

Kevin Camilo Díaz Monje

Directora
M.Sc Nancy Yaneth Gelvez García

Universidad Distrital Francisco José de Caldas
Maestría en Ciencias de la Información y las Comunicaciones
Énfasis en ingeniería de software
Bogotá D.C., Colombia
2025

DEDICATORIA

A mi directora de proyecto Nancy Y. Gelvez quien, con su guía, conocimiento, paciencia, motivación y sobre todo confianza logró que creyera en este proyecto y así convertirlo en una realidad.

A la docente Lilian Bejarano que más que una profesora fue una persona que siempre ha creído en mi como persona, como profesional y como estudiante.

A mis padres Rafael Díaz y Yasmin Monje y a mi hermana Daniela Díaz porque como todo lo que he conseguido es gracias a todo lo que me han brindado cariño, consejos y amor.

A mi amada Gisela María Díaz Gómez quien me vio múltiples veces a punto de desfallecer, pero me motivo para que yo no me diera por vencido, infinitamente agradecido por todo el amor brindado.

A Jeisson Piñeros, Daniel Hernández y Andrés Sánchez, mis amigos que estuvieron en esta larga travesía que juntos iniciamos y que logramos culminar, sin ustedes este reto no hubiera sido posible alcanzarlo.

A Juan Torres, Felipe Carrillo, Oscar Cárdenas, Juan Camiro Rodríguez y Camilo Ramírez que a lo largo de los estos años siempre me brindaron esa alegría que uno necesita para seguir.

Kevin Camilo Díaz Monje

AGRADECIMIENTOS

Agradecer es poder valorar hasta las cosas más sencillas, y por ello agradezco a todas las personas que me motivaron, que vieron en mi algo extraordinario, agradecido con cada una de las personas que me aportaron algo para continuar con este desafío.

Agradezco en específico a los docentes que estuvieron apoyando el proceso, al Docente Guillermo Guarnizo, al docente Alejandro Daza, al docente Paulo Gaona, al docente Carlos Montenegro y a los docentes que conforman el grupo de investigación GIIRA que siempre estuvieron pendientes del progreso de este proyecto.

Al docente Gerardo Alcides Muñoz Quiñones quien como jurado dio grandes aportes para la mejora del proyecto.

A mi jefe y gran amiga, Sonia Hernández que me acompañó durante todo el proceso del desarrollo de esta maestría.

Kevin Camilo Díaz Monje

Tabla de Contenido

Tabla de Contenido III

Listas de ilustraciones VI

Lista de Tablas IX

Lista de Ecuaciones IX

1 Definición del Proyecto 2

1.1 Introducción 2

1.2 Planteamiento del problema 4

1.3 Formulación del problema 6

1.4 Sistematización del problema 6

1.5 Objetivos 8

1.5.1 Objetivo General 8

1.5.2 Objetivos Específicos 8

1.6 Justificación 9

1.6.1 Justificación Teórica 9

1.6.2 Justificación Metodológica 11

1.6.3 Justificación Práctica 13

2 Marco Referencial 16

2.1 Introducción 16

2.2 Marco Teórico 17

2.2.1 Marco teórico enfocado a clasificación de imágenes: 17

2.2.2 Marco teórico enfocado a Multiview Learning: 18

2.2.3 Marco teórico enfocado al conflicto armado: 35

2.3 Marco Conceptual 40

3 Hipótesis 55
4 Metodología 57
4.1 Definición y contextualización 57
4.2 Modelamiento del Sistema Multi View Learning (MVL) 58
4.3 Modelamiento del sistema integrados de visualización 59
4.4 Pruebas, Análisis de resultados y conclusiones 60
5 Diseño del modelo MVL 62
5.1 Boceto inicial del diseño del Modelo 62
5.2 Definición de las capas del modelo 63
5.2.1 Vista de Redes Neuronales Convolucionales 65
5.2.2 Vista de Análisis de metadatos 65
5.2.3 Vista de Detección de Objetos 66
5.2.4 Vista de Retroalimentación de los usuarios 66
5.3 Modelo de Multiview Learning 67
6 Implementación del modelo 70
6.1 Introducción 70
6.2 Estructuración de la Base de datos 70
6.2.1 Almacenamiento de la información de las imágenes 71
6.2.2 Almacenamiento de información de usuarios 79
6.2.3 Almacenamiento de información de resultados 79
6.3 Arquitectura 80
6.4 Implementación de la vista de Redes Neuronales Convolucionales 81
6.5 Implementación de la vista de Análisis de metadatos 88
6.6 Implementación de la vista de Detección de Objetos 92

6.6.1 Estructuración y entrenamiento 92
6.6.2 Validación 94
6.7 Implementación del módulo de Multiview Learning 97
6.8 Implementación del sitio web de ejecución 100
6.9 Implementación de la vista de retroalimentación 102
7 Métricas y análisis de resultados 107
7.1 Métricas 107
7.1.1 Métrica de precisión 107
7.1.2 Métrica F1-Score 108
7.1.3 Métrica de satisfacción del usuario 109
7.2 Comparativa con el modelo MVMED+ 112
7.3 Análisis de resultados 115
8 Conclusiones 117
8.1 Conclusiones 117
8.2 Aporte al conocimiento 118
8.3 Recomendaciones para futuros trabajos 119
9 Bibliografía 121

Listas de ilustraciones

Ilustración 1: La exactitud del MVMED+ con diferentes números de vistas. Fuente: “Multi-view learning for visual violence recognition with maximum entropy discrimination and deep features discrimination and deep features” (Sun, Liu, & Mao, 2018) 20

Ilustración 2: Tabla de Performance Comparison os VSVC an other Baselne Methods. Fuente: (Zhang, Cheng, & Tian, 2019). 21

Ilustración 3: Muestra del proceso de multiview con IrMVL y MC. Fuente: “Low-Rank Multi-View Learning in Matrix Completion for Multi- Label Image Classification” (Liu, Luo, Tao, Xu, & Wen, 2015) 23

Ilustración 4: Multi-view CNN for 3D shape recognition. Fuente: Multi-view Convolutional Neural Networks for 3D Shape Recognition” (Su, Maji, Kalogerakis, & Miller, 2015) 23

Ilustración 5: Muestra del proceso de multiview learning active. Fuente: “A Deep Multiview Active Learning for Large-Scale Image Classification” (Tuozhong, Wenfeng, & Yuhong, 2020) 26

Ilustración 6: Muestra de la arquitectura MVMDCNN. Fuente: “Multi-View Convolutional Neural Networks for Mammographic Image Classification” (SUN, WANG, HU, XU, & Cui, 2019) 27

Ilustración 7: Muestra de Multiple View Oriented Matching Algorithm. Fuente: “Multiple View Oriented Matching Algorithm for People Reidentification” (García, Gardel, Bravo, & Lázaro, 2014) 29

Ilustración 8: Muestra de Entrenamiento y predicción del modelo de decodificación Neuronal. Fuente: “Reconstructing Perceived Images From Human Brain Activities With Bayesian Deep Multiview Learning” (Du, Du, Huang, & He, 2018) 31

Ilustración 9: (D-MVE-Hash) framework. Fuente: “Deep Multi-View Enhancement Hashing for Image Retrieval” (Yan, Gong, Wei, & Gao, 2020) 32

Ilustración 10: Muestra de la estructura del manejo de etiquetas Multimodal. Fuente: “Multimodal Transformer with Multi-View Visual Representation for Image Captioning” (Yu, Li, Yu, & Huang, 2020) 33

Ilustración 11: Muestra de la estructura de la arquitectura del MVMED. Fuente: “Multi-view Learning and Deep Learning for Microscopic Neuroblastoma Pathology Image Diagnosis” (Liu, Sun, & Yin, 2018) 34
Ilustración 12: Diagrama de secuencia del modelo de Clasificación. Fuente: Autor 63
Ilustración 13: Resultados de la prueba de validación de la técnica LBP. Fuente: Autor 64
Ilustración 14: Diagrama de componentes del sistema de Clasificación de Imágenes en conjunto con un sistema de Visualización. Fuente: Autor 67
Ilustración 15: Diagrama Entidad Relación de la base de datos utilizadas para el proyecto. Fuente: Autor 71
Ilustración 16: Muestra de la tabla de imágenes con el título, ubicación y fuerte de las imágenes. Fuente: Autor 72
Ilustración 17: Vista inicial del código del scrapper. Fuente: Autor 73
Ilustración 18: Muestra de la tabla de imágenes con solo imágenes obtenidas del scrapper. Fuente: Autor 73
Ilustración 19: Tabla de Metadato. Fuente: Autor 74
Ilustración 20: Tabla MetadatoImagen. Fuente: Autor 74
Ilustración 21: Visualización del código para la conexión y envió del resultado de cada vista. Fuente: Autor 75
Ilustración 22: Visualización de la estructura del repositorio de imágenes. Fuente: Autor 76
Ilustración 23: Imágenes del repositorio enfocadas al conflicto armado. Fuente: Autor ... 77
Ilustración 24: Tabla de ProcesoMultiview. Fuente: Autor 79
Ilustración 25: Tabla de ResultadosProcesosMultiview. Fuente: Autor 80
Ilustración 26: Visualización de los procesos de cada vista y de Multiview Dockerizados. Fuente: Autor 81
Ilustración 27: Visualización del código de preparación del dataset. Fuente: Autor 82
Ilustración 28: Vista de la jerarquía de carpetas para el manejo de las imágenes. Fuente: Autor 82
Ilustración 29: Visualización del código donde se estableció la red neuronal. Fuente: Autor 83

Ilustración 30: Resultado de la ejecución completa del proceso. Fuente: Autor 86
Ilustración 31: Visualización del diccionario de pesos. Fuente: Autor 88
Ilustración 32: Visualización del diccionario de conectores. Fuente: Autor 89
Ilustración 33: Nombres de los objetos a trabajar. Fuente: Autor 93
Ilustración 34: Resultado de un proceso de detección de objetos sobre una imagen. Fuente: Autor 95
Ilustración 35: Imagen de referencia para el ejemplo de detección de objetos. Fuente: Autor 96
Ilustración 36: Diagrama del funcionamiento del perceptrón. Fuente: Autor 97
Ilustración 37: Vista del código de entrenamiento para el perceptrón. Fuente: Autor 99
Ilustración 38: Diagramación de los resultados de cada una de las vistas. Fuente: Autor .. 99
Ilustración 39: Visualización de la aplicación web. Fuente: Autor 100
Ilustración 40: Vista de la sección para ejecución de procesos por vista del modelo. . Fuente: Autor 101
Ilustración 41: Imagen de presentación de la aplicación del sistema de visualización. Fuente: “DISEÑO DE UN METAMODELO DE SOFTWARE PARA EL DESARROLLO DE APLICACIONES INMERSIVAS” (Sánchez Cruz, 2025) 102
Ilustración 42: Imagen de la aplicación en la parte de calificación de la imagen. Fuente: “DISEÑO DE UN METAMODELO DE SOFTWARE PARA EL DESARROLLO DE APLICACIONES INMERSIVAS” (Sánchez Cruz, 2025) 103
Ilustración 43: Visualización del código de conexión para el consumo de las imágenes por parte del sistema de visualización. Fuente: Autor 103
Ilustración 44: Visualización del código para envió de la Calificación. Fuente: Autor 104
Ilustración 45; Código MVMED+ implementado para la comparación. Fuente: Autor 113
Ilustración 46: Resultado de la ejecución del modelo MVMED+ con el dataset del conflicto armado. Fuente: Autor 113

Lista de Tablas

Tabla 1: Ejemplos de la forma en la cual se analiza la descripción de una imagen. Fuente: Autor 90
Tabla 2: Tabla de resultados de clasificaciones Correctas. Fuente: Autor 108
Tabla 3: Tabla de preguntas al usuario. Fuente: Autor 110
Tabla 4: Tabla de resultados por pregunta y aspecto a valorar. Fuente: Autor 111
Tabla 5: Tabla de comparación de los modelos presentada en el artículo de comparación. Fuente: "Multi-view learning for visual violence recognition with maximum entropy discrimination and deep features discrimination and deep features" de (Sun, Liu, & Mao, 2018) 112
Tabla 6: Tabla de comparación entre resultados. Fuente: Autor 114

Lista de Ecuaciones

Ecuación 1: Estructura del vector de resultados de la capa de Redes Neuronales convolucionales. Fuente: Autor 86
Ecuación 2: Cálculo del resultado de análisis de metadato. Fuente: Autor 89
Ecuación 3: Estructura del vector de resultados de la capa de análisis de metadatos. Fuente: Autor 91
Ecuación 4: Cálculo del resultado de identificación de objetos. Fuente: Autor 96
Ecuación 5: Estructura del vector de resultados de la capa de Detección de Objetos. Fuente: Autor 96
Ecuación 6: Cálculo para el resultado del Multiview learning. Fuente: Autor 97
Ecuación 7: Cálculo para el resultado del Multiview learning con la cuarta vista. Fuente: Autor 105
Ecuación 8: Formula de la métrica de precisión. Fuente: "Multi-view learning for visual violence recognition with maximum entropy discrimination and deep features discrimination and deep features" de (Sun, Liu, & Mao, 2018) 107
Ecuación 9: Resultado de la métrica de precisión. Fuente: Autor 108

Ecuación 10: Fórmula de la métrica de F1-score. Fuente: “Multi-view learning for visual violence recognition with maximum entropy discrimination and deep features discrimination and deep features” de (Sun, Liu, & Mao, 2018) 108
Ecuación 11: Fórmula de la métrica de precisión para la métrica F1-score. Fuente: Autor 108
Ecuación 12: Fórmula de la métrica de Recall para la métrica F1-score. Fuente: Autor .. 109
Ecuación 13: Resultado de la fórmula de la métrica de precisión para la métrica F1-score. Fuente: Autor 109
Ecuación 14: Fórmula de la métrica de Recall para la métrica F1-score. Fuente: Autor .. 109
Ecuación 15: Fórmula de la métrica de F1-score. Fuente: Autor 109

Capítulo 1

Definición del Proyecto

Poder crear algo de la propia imaginación, pensarlo, trabajarlo, soñarlo, idealizarlo, construirlo, volverlo a plantear, fracasar y volver a crear es parte de la naturaleza de una mente creativa, y el hecho de definir un proyecto es tan solo el primer paso de una larga travesía con uno mismo.

Fuente: Autor

1 Definición del Proyecto

1.1 Introducción

La clasificación de imágenes es un área de las ciencias de la computación donde más se han hecho avances en investigación durante los últimos años. Esta disciplina se enfoca en el desarrollo de algoritmos y técnicas que permiten la interpretación automática de contenido visual, posibilitando tareas como la segmentación, el reconocimiento de objetos y la clasificación de imágenes. Como se presenta en el artículo titulado *"Advances in Computer Vision-Based Civil Infrastructure Inspection and Monitoring"* (Spencer, Hoskere, & Narazaki, 2019) "La visión por computadora es un campo científico interdisciplinario que se ocupa de la extracción automática de información útil de los datos de imágenes para comprender o representar el mundo físico subyacente, ya sea cualitativa o cuantitativamente."

Uno de los desafíos actuales en este campo es la clasificación eficiente del creciente volumen de contenido visual en plataformas digitales y entornos inmersivos, como la realidad aumentada (AR) y la realidad virtual (VR). En estos contextos, la personalización del contenido juega un papel fundamental en la experiencia del usuario. Sin embargo, muchos sistemas de clasificación actuales se basan en un único criterio de análisis, determinado por el algoritmo o los datos de entrenamiento, lo que puede resultar en resultados limitados o sesgados.

Enmarcado en este contexto y viendo la necesidad de tener una clasificación que tenga en cuenta diversos factores, el presente trabajo propone un enfoque basado en **Multiview Learning (ML)**, una técnica de aprendizaje automático que permite integrar múltiples representaciones o vistas de los datos. Este enfoque resulta especialmente útil para mejorar la precisión en la clasificación de imágenes, ya que permite considerar diversos criterios simultáneamente. En particular, este proyecto se enfoca en la clasificación de imágenes relacionadas con contextos culturales, como aquellas asociadas al conflicto armado en Colombia, con el objetivo de ofrecer una categorización más contextualizada y significativa para los usuarios.

En resumen, el presente trabajo propone y desarrolla un modelo de clasificación de imágenes basado en Multiview Learning (MVL), cuyo alcance se centra en la generación y validación del modelo para la mejora en la precisión y la pertinencia de los sistemas de visualización de contenido. El proyecto integra diversas vistas de las imágenes Redes neuronales convolucionales, Análisis de metadatos, Detección de objetos y Retroalimentación de los usuarios, con el fin de ofrecer clasificaciones más robustas, contextualizadas y significativas.

El caso de estudio se centra en imágenes vinculadas al conflicto armado en Colombia, precisamente por la complejidad cultural, semántica e histórica que representa este tipo de material. Esta elección no solo plantea un reto técnico para el modelo, sino que también aporta un valor social al facilitar interpretaciones más precisas de contenidos sensibles en entornos inmersivos. La iniciativa de este caso de estudio y sobre todo su forma de abordarlo con técnica de aprendizaje automático se sustenta por investigaciones que han empleado técnicas avanzadas como el aprendizaje automático para analizar, entender y predecir fenómenos asociados a la violencia armada en el país tal como se evidencia en el articulo “Predicting police and military violence: evidence from Colombia and Mexico using machine learning models” (Gelvez, 2025), en donde se abordan temas de ejecuciones extrajudiciales en Colombia en modelos predictivos para identificar patrones de violencia.

Asimismo, este trabajo se enmarca en el proyecto de la tesis doctoral *"Modelo computacional cognitivo para la predicción de preferencias de usuario, basado en el análisis de emociones y preclasificación de contenido, orientado a la visualización de datos"*, de la directora NANCY GELVEZ GARCIA.

Además de articularse con el trabajo de tesis de Maestría del ingeniero Andrés Felipe Sánchez Cruz titulado “DISEÑO DE UN METAMODELO DE SOFTWARE PARA EL DESARROLLO DE APLICACIONES INMERSIVAS” (Sánchez Cruz, 2025) el cual se usa como la herramienta para el trabajo en conjunto para un sistema de visualización completo de imágenes.

1.2 Planteamiento del problema

En la actualidad, el acceso a contenido digital es masivo y diverso, presentándose en múltiples formatos como imágenes, audio, video y contenido interactivo. Este contenido es presentado así mismo a través de diversas plataformas, como redes sociales, aplicaciones móviles y entornos inmersivos (realidad virtual, realidad aumentada, etc.), donde la experiencia del usuario depende no solo de la calidad visual, sino también de la pertinencia del contenido presentado.

La necesidad de presentar contenido a las personas que mejore su experiencia ha llevado al desarrollo de software a expandir sus capacidades implementando sistemas de recomendación que sugieren contenido basado en patrones de consumo. Si bien estos sistemas priorizan los gustos de las personas no llegan a realizar un análisis del contenido visual en profundidad antes de su presentación a los usuarios. Esta falta de una clasificación contextualizada y personalizada de las imágenes limita la capacidad de los sistemas de visualización para ofrecer experiencias relevantes, ricas y alineadas con los intereses o el conocimiento del usuario. En consecuencia, los usuarios reciben contenido que puede resultar genérico, poco contextualizado o incluso confuso, especialmente cuando las imágenes tratan temas complejos, como el conflicto armado, que requieren interpretación semántica y sensibilidad social y cultural.

Se puede definir como un sistema de visualización los diversos entornos inmersivos y aplicaciones de visualización (como realidad virtual, aumentada o plataformas interactivas), en donde el contenido visual juega un papel clave en la experiencia del usuario estos sistemas de visualización se centran en aspectos técnicos como la calidad gráfica, la interactividad y la facilidad de uso, pero no abordan de manera adecuada la riqueza semántica de las imágenes ni la adaptación del contenido al contexto del usuario ni las múltiples características que una imagen puede contener. Esto se debe a que los métodos tradicionales de clasificación, basados en etiquetas predefinidas o enfoques unidimensionales, resultan insuficientes para capturar las múltiples perspectivas y

significados que una imagen puede contener (Manning, Raghavan, & Schütze, 2009). Esto afecta la capacidad del sistema para presentar contenido visual de manera efectiva, tanto a nivel informativo como emocional, reduciendo la personalización, la diversidad y el impacto de la experiencia. (Recommender Systems: The Textbook, 2016).

De seguir esta situación, el contenido que se está consumiendo por parte de los usuarios de estas plataformas se verá restringido a el contenido preestablecido para dicha visualización, sin generar un valor agregado o un contenido diverso ya sea en conocimiento o interés para cada uno de los usuarios en particular, siendo aspectos que mejoran la experiencia del usuario. Dado este escenario, se entiende que las aplicaciones requieren un manejo del contenido previo que vaya un poco más allá de una simple clasificación y que de esta manera brinda imágenes que sean relevantes para un contexto especifico y así evitando el decaimiento del interés de los usuarios en los sistemas de visualización. De no darse un mejoramiento en la clasificación del contenido se ocasiona que esos entornos sean condenados a la obsolescencia temprana o falta de uso ya que no pueden retener a los usuarios por no generar interés. Como se muestra en el estudio "*2018 App Retention Benchmarks [Data Study]"* (Appflyer, 2018)

Para abordar esta problemática, se propone implementar la técnica de Multiview Learning en un sistema de clasificación y que este sea parte de los sistemas de visualización en contextos culturales complejos —como el caso de estudio centrado en imágenes del conflicto armado en Colombia— para potenciar significativamente la calidad de la experiencia del usuario. Al lograr una mejor contextualización del contenido visual, se amplía la diversidad informativa, se favorece la apropiación del conocimiento y se fortalece la conexión emocional del usuario con la aplicación.

Con lo anteriormente presentado se establece que el problema central radica en la falta de una clasificación contextualizada, semántica y adaptativa de las imágenes en sistemas de visualización inmersiva, lo que reduce la relevancia del contenido presentado, afecta la retención del usuario y limita el potencial de estas plataformas. Superar esta limitación

exige un nuevo enfoque basado en el análisis de múltiples vistas sobre las imágenes, que permita ofrecer experiencias personalizadas, significativas y contextualmente pertinentes.

1.3 Formulación del problema

Ya con el planteamiento del problema e identificando el propósito del proyecto la pregunta que enmarca la situación es:

¿Cómo mejorar la clasificación de imágenes mediante Multiview Learning para lograr una presentación contextualizada y significativa en entornos inmersivos?

1.4 Sistematización del problema

Para poder dar soporte al problema planteado se plantean algunos interrogantes encaminados a generar tal solución, estas preguntas son:

- ¿Cuáles son las características intrínsecas (como objetos, patrones contextuales y atributos semánticos), que deben considerarse en las imágenes para ser tratadas como vistas individuales dentro de un modelo de clasificación basado en *Multiview Learning*?
- ¿Cómo pueden integrarse técnicas de procesamiento de imágenes con algoritmos de *Multiview Learning* para mejorar la precisión en la clasificación de contenido visual?
- ¿De qué manera puede incorporarse la retroalimentación de los usuarios al modelo de *Multiview Learning* para refinar y adaptar continuamente la clasificación de imágenes, permitiendo mejorar la experiencia de cada usuario?

- ¿Qué criterios deben aplicarse para evaluar la efectividad de la clasificación en términos de pertinencia e impacto en la experiencia del usuario dentro de sistemas de visualización?

1.5 Objetivos

1.5.1 Objetivo General

Diseñar un modelo de clasificación de imágenes basado en Multiview Learning que permita mejorar la precisión en la clasificación y la pertinencia del contenido presentado en sistemas de visualización.

1.5.2 Objetivos Específicos

- Identificar las características relevantes de las imágenes que pueden ser utilizadas para definir múltiples vistas en el proceso de clasificación dentro del modelo de *Multiview Learning*.
- Diseñar un modelo de clasificación basado en Multiview Learning, integrando técnicas de procesamiento de imágenes y análisis de texto, para mejorar la clasificación de contenido visual.
- Implementar un prototipo de software que integre la retroalimentación de las personas junto con el modelo propuesto, permitiendo la clasificación automatizada de contenido visual para entornos inmersivos y sistemas de visualización.
- Establecer criterios de evaluación para medir la efectividad de la clasificación en términos de pertinencia del contenido e impacto en la experiencia del usuario en entornos inmersivos.

1.6 Justificación

Esta investigación tiene como base el identificar y analizar las características intrínsecas presentes en las imágenes que serán procesadas, con el fin de desarrollar un sistema de clasificación de contenido que tenga en consideración esas características al momento de realizar la clasificación. Dicho sistema busca poder brindar una mejora en la clasificación de contenido y por ende en la experiencia del usuario en sistemas de visualización al presentar un contenido pertinente al contexto.

1.6.1 Justificación Teórica

El presente trabajo se fundamenta en la necesidad de fortalecer los procesos de clasificación de imágenes en contextos donde los contenidos visuales exigen un análisis más profundo que el que permiten los modelos tradicionales basados en etiquetas predefinidas o criterios unidimensionales. Para abordar este desafío, se parte de los avances actuales en el tratamiento de imágenes, como emplear técnicas de procesamiento de imágenes y enfoques innovadores como *Multiview Learning* (ML), los cuales permiten abordar la caracterización de imágenes desde múltiples perspectivas extrayendo características más representativas y mejorando su clasificación.

Para abordar este trabajo es importante dar la relevancia a los sistemas de clasificación de imágenes, que más que una tarea consiste en una serie de procesos para organizar, interpretar y presentar información visual de forma efectiva, especialmente en contextos donde las imágenes contienen significados complejos o contextuales. El hecho de que no se genere una clasificación adecuada, las imágenes pueden ser mostradas fuera de contexto, malinterpretadas o simplemente ignoradas por los usuarios, lo cual limita el valor informativo, educativo o narrativo de los sistemas de visualización y el impacto emocional que puede generar en las personas.

Considerando la relevancia de los sistemas de clasificación de imágenes, es importante señalar que muchos de los enfoques actuales aún se basan en etiquetas predefinidas o en modelos entrenados sobre características visuales superficiales. Estos sistemas, por lo general, adoptan una perspectiva de vista única, sin contemplar las posibles relaciones

contextuales en torno a la creación o el significado de la imagen. Esta limitación restringe su capacidad para captar la complejidad semántica y las múltiples dimensiones culturales o simbólicas que una imagen puede contener.

Desde una perspectiva teórica, esta investigación se sustenta en dos pilares fundamentales: el procesamiento de imágenes y el aprendizaje automático basado en múltiples vistas, donde se busca utilizar algoritmos/técnicas de manejo de imágenes para establecer características en común entre cada una de ellas y posteriormente aplicar Multiview learning para generar las vistas. Cada vista busca capturar los elementos más representativos de una imagen, resaltando los atributos distintivos que caracterizan a cada conjunto. Para ello, se emplean algoritmos y técnicas de procesamiento de imágenes, cuyo propósito es identificar patrones comunes entre las muestras. Una vez establecidas estas características compartidas, se aplica el enfoque de Multiview Learning para generar representaciones que sinteticen la información más relevante de forma estructurada y significativa.

Esta investigación busca partir de un componente teórico basado en el tratamiento de imágenes mediante modelos existentes, utilizar las características intrínsecas de las imágenes de contenido cultural, para realizar una clasificación de dichas imágenes, de tal forma que la información de la clasificación pueda ser utilizada al momento de querer presentar contenido en una técnica de visualización.

Esto, junto con la retroalimentación proveniente del sistema de visualización, permite comparar los resultados obtenidos inicialmente e integrarlos, lo cual contribuye a una mejor clasificación por parte del sistema de Multiview Learning y repercute directamente en la calidad de los sistemas de visualización. En síntesis, esta investigación busca combinar fundamentos teóricos, técnicas de procesamiento de imágenes y sistemas de clasificación de múltiples vistas, con el fin de potenciar el manejo y la representación del contenido en entornos tecnológicos innovadores.

1.6.2 Justificación Metodológica

Desde un enfoque metodológico, el presente trabajo propone el diseño de un modelo de clasificación de imágenes mediante la implementación de algoritmos de reconocimiento y procesamiento de imágenes. Para ello, se plantea la incorporación del concepto de *Multiview Learning* (ML), que permite integrar el análisis de las imágenes desde múltiples perspectivas, en donde se identifiquen dimensiones semánticas, visuales y de retroalimentación del usuario en un mismo modelo.

Este planteamiento de una nueva forma de clasificación es necesario porque los enfoques convencionales no logran capturar adecuadamente la complejidad visual y semántica que puede contener muchas imágenes, especialmente en contextos reales que sean muy específicos como lo pueden ser contextos narrativos, culturales, históricos, entre otros.

Desde una perspectiva que busca justificar el uso del Multiview Learning como eje central del proyecto, esta técnica se considera adecuada porque permite analizar una imagen desde múltiples vistas que se complementan entre sí. Esto enriquece significativamente el proceso de clasificación, al brindar la posibilidad de combinar diversos algoritmos y enfoques, lo cual lo hace más robusto, preciso, adaptativo y, sobre todo, más completo para escenarios en los que la imagen, por sí sola, requiere interpretaciones más complejas.

La metodología se fundamenta en el desarrollo progresivo de algoritmos de clasificación. En una primera etapa, se implementarán algoritmos de procesamiento de imágenes, cuyos resultados serán evaluados y ajustados de forma iterativa para lograr una clasificación precisa y coherente con el contexto. Inicialmente, el análisis de los patrones identificados se realizará de manera manual; no obstante, a medida que los resultados se aproximen a lo esperado, se avanzará hacia la automatización del proceso. Paralelamente, y conforme el modelo se refine, se incorporarán métodos estadísticos exploratorios que optimicen la interpretación de los resultados, garantizando una clasificación eficiente, adaptativa y contextualizada.

Adicionalmente, se incorporará una capa de retroalimentación al modelo de Multiview Learning, permitiendo que los usuarios puedan contribuir en la mejora de la precisión al

clasificar, a través de su interacción con los sistemas de visualización. Esta retroalimentación será integrada en el proceso de aprendizaje del modelo, refinando continuamente los resultados y asegurando que la clasificación se adapte de manera dinámica a diferentes contextos.

En resumen, esta metodología combina el uso de algoritmos avanzados de reconocimiento de imágenes, la automatización progresiva del análisis de datos y la retroalimentación basada en la experiencia del usuario, con el fin de desarrollar un sistema de clasificación de imágenes más preciso y adaptable, que pueda ser aplicado en diversos contextos de visualización y representación de contenido.

1.6.3 Justificación Práctica

A nivel práctico, este trabajo de investigación busca dar respuesta a una problemática concreta: la baja contextualización y pertinencia del contenido visual en los sistemas de visualización. Al centrarse en imágenes, el modelo propuesto basado en Multiview Learning tiene un alto potencial e impacto no solo en la mejora técnica de los sistemas de clasificación y, por consiguiente, en los de visualización, sino también en la construcción de una experiencia de usuario más significativa y coherente con el contexto sobre el cual se trabaja.

Este enfoque permite analizar y organizar contenido visual de manera más precisa, ya que aprovecha múltiples perspectivas para extraer características relevantes antes de su presentación. Al mejorar la clasificación, se incrementa la relevancia del contenido mostrado, facilitando una experiencia más adaptada a las necesidades y preferencias de los usuarios. Esto es crucial, ya que asegura que el contenido que los usuarios reciben ha pasado por un proceso de evaluación y clasificación previa.

Además, esta metodología permite identificar áreas de mejora basadas en las calificaciones y percepciones de los usuarios, lo que incrementa la calidad y relevancia del contenido presentado. Su aplicación no se limita a un solo contexto, sino que es extensible a diversos entornos productivos, ya que el análisis de imágenes puede revelar relaciones y patrones relevantes entre conjuntos de imágenes aparentemente diferentes.

Este modelo no solo es valioso para mejorar la presentación de imágenes en aplicaciones interactivas, sino que también ofrece una base flexible para futuras investigaciones y desarrollos tecnológicos. Su implementación puede contribuir a la evolución de herramientas de clasificación visual en diferentes industrias, impulsando la adopción de inteligencia artificial en la gestión de contenido visual. Además, su integración con infraestructuras tecnológicas existentes puede generar beneficios económicos al reducir costos operativos y mejorar la eficiencia en el manejo de grandes volúmenes de datos visuales.

Asimismo, el desarrollo de este modelo de Multiview Learning en sistemas de clasificación de imágenes representa un avance significativo para su implementación en sistemas de visualización, aplicable a múltiples ámbitos como museos digitales, entornos educativos en realidad aumentada o virtual, y plataformas culturales interactivas. Este enfoque permite ofrecer una experiencia más profunda y personalizada. Su capacidad para mejorar la clasificación —en términos de organización, relevancia y pertinencia del contenido visual— no solo transforma la experiencia del usuario, sino que también abre nuevas oportunidades para la innovación y el desarrollo tecnológico en diversos sectores.

Capítulo 2

Marco Referencial

¿Cómo investigar sin querer crear?,
¿Cómo crear sin saber hacer?,
¿Cómo hacer sin conocer?,
¿Cómo conocer sin aprender?,
¿Cómo aprender sin leer?,
¿Cómo leer sin placer?.

Al final Investigar es un placer

Fuente: Autor

2 Marco Referencial

2.1 Introducción

A partir del planteamiento del problema, resulta fundamental llevar a cabo una revisión exhaustiva de la literatura académica relacionada con el Multiview learning, la clasificación de imágenes y otros temas afines que sustentan de forma teórica y práctica esta investigación. Este análisis permitirá identificar las problemáticas existentes, los escenarios de aplicación, los avances recientes y las limitaciones de los enfoques actuales. Asimismo, la revisión previa facilitará la delimitación del alcance y los criterios metodológicos que guiarán este proyecto, asegurando una base sólida para su desarrollo.

Como parte de la revisión de trabajos previos, se tuvo presente el articulo denominado *"Estudio y estructuración bibliográfica del estado del arte del uso de multiview learning en la clasificación de imágenes"* (Diaz Monje, 2022) , donde se expone una metodología de revisión bibliográfica que incluye la definición y delimitación del problema —el uso de Multiview Learning para la clasificación de imágenes—, así como el análisis de múltiples fuentes teóricas relevantes. Esta revisión permitió comprender los fundamentos conceptuales, el estado del arte y la evolución de la temática, insumos esenciales para el desarrollo del presente proyecto.

Adicional a este trabajo se hace también alusión a diversas técnicas, tecnologías y algoritmos que se utilizaron para la realización del proyecto.

2.2 Marco Teórico

Al momento de realizar la lectura de cada uno de los artículos se buscó rescatar la información más relevante de cada uno de ellos con la idea de hacer un breve resumen de su contenido y hacer referencia a dichos trabajos.

2.2.1 Marco teórico enfocado a clasificación de imágenes:

Se inició con una búsqueda enfocada a la clasificación de imágenes partiendo del concepto que se maneja entro de los campos de la visión por computadora y el aprendizaje automático en donde se refiere al proceso de asignar una etiqueta o categoría a una imagen, basándose en su contenido intrínseco de lo que quiere representar la imagen.

En esta búsqueda se encuentran diversas ramas de estudio como lo es el *Deep learning* o aprendizaje profundo, donde las Redes Neuronales Convolucionales (CNN) son el estándar en la clasificación de imágenes debido a su capacidad para aprender características de alto nivel directamente desde los datos ya que utilizan capas convolucionales para extraer características de las imágenes, eliminando la necesidad de extracción manual de elementos intrínsecos de cada imagen. Un ejemplo de ello es la revisión realizada en el artículo "*Image classification based deep learning: A Review*" (Mohamed, Hemeida, & Hassan, 2022) el cual brinda una base teórica bastante enfocada a este tema de estudio. También se ha implementado en la clasificación de imágenes complejas tal como se ve en el artículo denominado "*ImageNet Classification with Deep Convolutional*" (Krizhevsky, Sutskever, & Hinton, 2012) en donde se ha hecho la clasificación en 5 posibles etiquetas para un conjunto de entrenamiento.

Otro concepto base es *¿cómo se hace la evaluación de los modelos de clasificación de imágenes?* ya que es esencial para garantizar su efectividad. Las métricas más comunes incluyen la precisión, el recall, la F1-score y la matriz de confusión. Estas métricas permiten analizar el rendimiento del modelo y ajustar los parámetros según sea necesario. Tal como se trata en el artículo "*A systematic analysis of performance measures for classification tasks*" (Sokolova & Lapalme, 2009) en el cual se estudia cómo se utilizan las métricas en

procesos de clasificación de imágenes en entornos de clasificación binaria, multicase y multietiqueta.

2.2.2 Marco teórico enfocado a Multiview Learning:

Debido a que Multiview learning (de aquí en adelante MVL) es una forma de implementación del machine learning que está teniendo su auge es importante conocer como se ha venido trabajando en el ámbito investigativo, uno de los primeros artículos que cabe resaltar es titulado como "*Multi-view learning for visual violence recognition with maximum entropy discrimination and deep features.*" (Sun, Liu, & Mao, 2018), el cual tiene gran relevancia para este proyecto, ya que brinda una percepción clara de lo que se puede hacer a nivel de tratamiento de imágenes, en este trabajo se entrena un algoritmo de reconocimiento de imágenes de índole violento para que mediante el uso de Deep learning estas imágenes pudieran ser analizadas y estratificadas por la violencia que contenga la imagen para así saber si es posible censurar el contenido o no.

Una base fundamental para comprender los procesos del Multiview Learning fue la clasificación de los algoritmos utilizados en esta técnica, organizándolos en tres categorías según su forma de entrenamiento. Estas categorías son:

- Algoritmos de estilo de co-entrenamiento
- Algoritmos de estilo de co-regularización
- Algoritmos de estilo de consistencia de márgenes

El trabajo titulado "*Multi-view learning overview: Recent progress and new challenges and new challenges.*" por (Zhao, Xijiong, Xin, & Shiliang, 2017) , se presentan algunos algoritmos para cada una de las categorías y su respectivo funcionamiento siendo un punto de referencia clave al ser un artículo de revisión bibliográfica enfocado al multiview.

Un artículo interesante el cual sirve como referencia para el manejo de imágenes es el artículo "*Multiview discriminative marginal metric learning for makeup face verification*" de (Zhang, Shum, Liu, Guo, & Shao, 2019) ya que toma como referencia algunos algoritmos de

bajo nivel como los son HOG, SIFT, TPLBP, LBP Y FBD, con el fin de determinar diferencias entre dos imágenes (una imagen con maquille y otra sin el maquillaje, esto en la misma persona), posteriormente aplican el principio básico de Multiview discriminative marginal metric learning (MDMML) para aplicar el modelo de clasificación.

Al realizar este estudio se encontraron algunos algoritmos interesantes y relevantes, los cuales son: algoritmos de bajo nivel, sobre los cuales también se soporta otro algoritmo de MVL llamado multi-view maximum entropy discriminant (MVMED+) basado en: model for learning with different numbers of views. o en español (modelo discriminante de máxima entropía de múltiples vistas (MVMED+)); este fue trabajado en el artículo "*Multi-view learning for visual violence recognition with maximum entropy discrimination and deep features discrimination and deep features*" de (Sun, Liu, & Mao, 2018), en el cual se trabaja bajo el objetivo de lograr reconocer imágenes que representan violencia física, para ello se soportan en los algoritmos de bajo nivel que luego se unirán con MVL, estos algoritmos son DSIFT Feature,HOG Features y LBP Features. Posteriormente, se introduce de forma matemática y algorítmica el concepto de la discriminación por máxima entropía en donde a grandes rasgos se trata de optimizar una función objetivo en la cual se incluyen los resultados de cada uno de los algoritmos de bajo nivel trabajados previamente. Adicionalmente establecer los valores requeridos para un SMO (Sequential minimal optimization). Todo esto en conjunción logra establecer unos resultados para un conjunto de imágenes en comparación con las imágenes tratadas de forma individual o en combinaciones pares de las vistas como se muestra en la Ilustración 1 del artículo "*Multi-view learning for visual violence recognition with maximum entropy discrimination and deep features discrimination and deep features*" (Sun, Liu, & Mao, 2018)

Table 1
The accuracy of MVMED$^+$ with different number of views.

Views	Life	Scene	Physical
view1	77.82 ± 0.47	89.94 ± 0.62	88.35 ± 0.70
view2	78.81 ± 0.45	92.38 ± 0.52	91.70 ± 0.47
view3	85.99 ± 0.57	92.26 ± 0.47	88.74 ± 0.95
view1 + view2	80.29 ± 0.63	94.06 ± 0.54	92.58 ± 0.63
view1 + view3	85.56 ± 0.71	93.07 ± 0.36	90.60 ± 0.60
view2 + view3	86.07 ± 0.46	94.29 ± 0.63	92.52 ± 0.70
view1 + view2 + view3	$\mathbf{86.12 \pm 0.42}$	$\mathbf{95.10 \pm 0.52}$	$\mathbf{92.60 \pm 0.41}$

Ilustración 1: La exactitud del MVMED+ con diferentes números de vistas. Fuente: "Multi-view learning for visual violence recognition with maximum entropy discrimination and deep features discrimination and deep features" (Sun, Liu, & Mao, 2018)

Otro trabajo importante es el artículo denominado "*A Survey on Multi-view Learning*" (Xu, Dacheng, & Xu, 2013) en el cual se presentan los dos principios de relación entre las múltiples vistas, los cuales son:

- El principio de consenso: el cual tiene como objetivo, manejar la información de las diferentes vistas y maximizar el "consenso" o la igualdad de la información entre las vistas. Esto muestra la conexión entre estas vistas bajo algunas superposiciones y leves desigualdades que puedan existir al maximizar la correlación de la información. Esto también se puede aplicar utilizando una minimización dependiendo del caso en concreto.
- El principio de complemento: Este principio se basa en que cada vista puede contener información que otra vista no puede contener mostrando diferente información sobre un mismo objeto de estudio en capas que de por sí manejan información que en esencia no es igual. Una forma de manejar este concepto es utilizar múltiples clasificadores para extraer información complementaria y que de forma paralela se puedan entrenar para hallar comportamientos que unan la información pero que a su vez den el complemento esperado.

Continuando con la revisión del artículo "*A Survey on Multi-view Learning*" (Xu, Dacheng, & Xu, 2013) se plantea el proceso de creación de uso básico de las técnicas de MVL, partiendo

de la manera en la que se deben crear las vistas, cómo se deberían combinar, manejando técnicas de co-entrenamiento y estableciendo algunos algoritmos de múltiple aprendizaje que se han trabajado a través de la evolución de MVL, como lo son Multiple Kernel Learning (MKL) y el Subspace Learning-based Approaches que a su vez exploran todos los algoritmos y técnicas que cada uno de estos conllevan.

En el artículo titulado "*Multi-View Image Classification With Visual, Semantic and View Consistency*" (Zhang, Cheng, & Tian, 2019) se trabaja la idea de consenso entre las múltiples vistas, esto implica que se busca que entre las vistas que se están manejado existen características que sean iguales o semejantes, para ello lo que realizaron fue aplicar cada una de las coherencias en vistas separadas y posteriormente utilizar clasificadores pre entrenados para combinar las coherencias, esto con un método denominado "Linear Multi-View Combination", posteriormente comienzan a experimentar con varios datasets para ver su comportamiento, y objetos que clasifica; también para realizar una comparación en desempeño conforme a otros métodos que se usaron como su línea base de comparación lo anterior se puede observar en la Ilustración 2 tomada del articulo "*Multi-View Image Classification With Visual, Semantic and View Consistency*" (Zhang, Cheng, & Tian, 2019).

TABLE V

PERFORMANCE COMPARISONS OF VSVC AND OTHER BASELINE METHODS ON THE CUB-200-2011 DATASET. BB: BOUNDING BOX

Methods	BB	Performance (%)
FC-VGG [45]	no	70.4
bilinear CNN [61]	no	84.1
LRBP[62]	no	84.2
PR-CNN [16]	yes	73.5
Triplet-A [63]	yes	80.7
STN [64]	no	84.1
BoostCNN [65]	no	86.2
VGG-D [66]	yes	82.0
RBF [73]	no	79.0
AGAL [74]	yes	85.5
VSVC(average)	no	87.5
VSVC	no	87.9

Ilustración 2: Tabla de Performance Comparison os VSVC an other Baselne Methods. Fuente: (Zhang, Cheng, & Tian, 2019).

Un artículo importante para tener en cuenta por establecer un marco de aprendizaje semi supervisado es el articulo "*Multi-view semi-supervised learning for image classification*" (Zhu, Sun, & Jin, 2016) en el cual toman la información de imágenes que están pre etiquetadas; se establece un sistema de etiquetado de imágenes para entrenar y clasificar las imágenes de forma independiente y posteriormente con estas vistas se usa un principio de mínima entropía para establecer la categoría de la etiqueta que debería llevar las imágenes y así poder clasificar la imágen de forma certera. Algo similar se ve en el artículo de la conferencia de la IEEE ICME del 2009 llamado "*Multi-view multi-label active learning for image classification*" (Zhang X. , Cheng, Xu, Lu, & Ma, 2009) el cual realiza un aprendizaje activo de múltiples vistas estableciendo una serie de incertidumbres en las vistas para posteriormente fusionarse a partir de una muestra-etiquetada que se tiene como base.

En el campo del Multiview Learning, es común encontrar propuestas orientadas a mejorar los algoritmos ya existentes, especialmente frente a los desafíos relacionados con la complejidad computacional y los tiempos de clasificación y procesamiento. Un artículo que habla de ello es "*Low-Rank Multi-View Learning in Matrix Completion for Multi- Label Image Classification*" (Liu, Luo, Tao, Xu & Wen, 2015), donde se propone el uso de la compleción de matrices como método para la clasificación multietiqueta. En este enfoque, se emplean algoritmos de bajo nivel como DSIFT, TAG y GIST, los cuales permiten obtener múltiples vistas de características de las imágenes. Posteriormente, se aplica el modelo denominado *Low-Rank Multi-View Matrix Completion (lrMMC)*, que utiliza la información recopilada para realizar la clasificación mediante la compleción matricial. Este proceso se ejemplifica en la Ilustración 3, tomada del artículo mencionado. (Liu, Luo, Tao, Xu, & Wen, 2015)

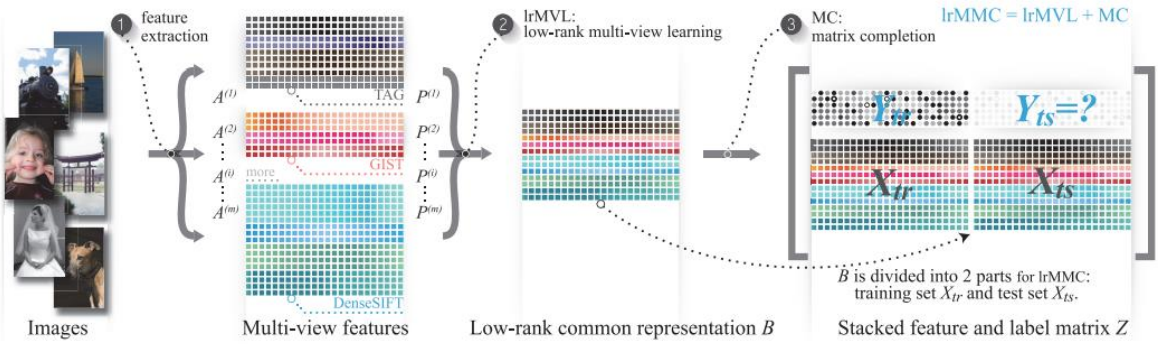

Ilustración 3: Muestra del proceso de multiview con lrMVL y MC. Fuente: "Low-Rank Multi-View Learning in Matrix Completion for Multi- Label Image Classification" (Liu, Luo, Tao, Xu, & Wen, 2015)

Un desafío que se presenta en la literatura consiste en el manejo de imágenes en 2D que representan objetos en tercera dimensión (3D) para lo cual se ha trabajado el MVL apoyado en técnicas de redes neuronales convolucionales, esto se plasma en el trabajo titulado "*Multi-view Convolutional Neural Networks for 3D Shape Recognition*" (Su, Maji, Kalogerakis, & Miller, 2015), en este articulo trabajan las imágenes de perspectivas de una objeto de forma separada, cada una de estas imágenes entra a una red neuronal, el resultado de cada una de ellas se agrupa y posteriormente se pasa la información a una red neuronal convolucional, para lograr identificar que estas vistas coinciden con un objeto en específico (una imagen de la figura 3D que está relacionada con múltiples imágenes 2D), esto se puede ver representado en la Ilustración 4 tomada del artículo "*Multi-view Convolutional Neural Networks for 3D Shape Recognition*" (Su, Maji, Kalogerakis, & Miller, 2015).

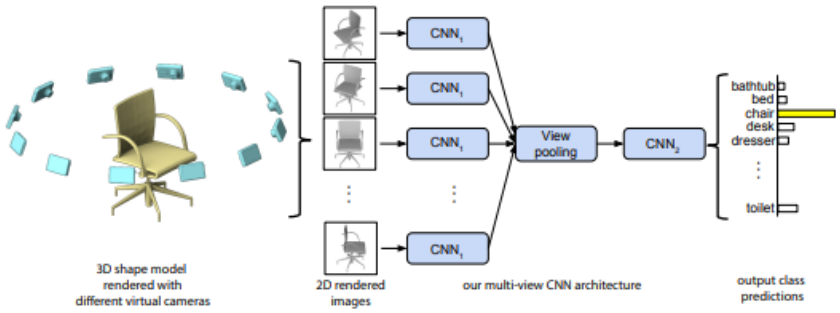

Ilustración 4: Multi-view CNN for 3D shape recognition. Fuente: Multi-view Convolutional Neural Networks for 3D Shape Recognition" (Su, Maji, Kalogerakis, & Miller, 2015)

Otro gran desafío que se trabaja en el MVL costa de buscar formas de mejorar el rendimiento de las clasificaciones, esto ha generado que se integren procesos con ramas de la inteligencia artificial como lo son las redes neuronales y más específicamente las redes neuronales convolucionales, esto se muestra en el artículo "*Multi-view classification with convolutional neural networks*" (Seeland & Mäder, 2020), en este trabajo se es utilizan las redes neuronales convolucionales para extraer la información de los objetos (Plantas insectos y autos) para así lograr implementar 3 formas diferentes de fusionar las múltiples vistas, estas son (1) fusionar mapas de características convolucionales a diferentes profundidades de red (Early fusion); (2) fusión de representaciones latentes de cuellos de botella antes de la clasificación (Late fusion); y (3) fusión de puntuación (Score fusion). De esta forma estiman cual es la mejor forma de combinar estas múltiples vistas para los entornos de dominio planteados.

Existe una gran variedad de algoritmos orientados a diferentes partes del proceso de MVL. Una de estas etapas clave es la fusión, concatenación o correlación de las múltiples vistas generadas a partir de los métodos establecidos en cada investigación. La integración de estas vistas se realiza en muchos casos, mediante redes neuronales. En este nuevo artículo llamado "*Deep Multi-View Learning using Neuron-Wise Correlation-Maximizing Regularizers*" (Jia, Lin, Tan, & Tao, 2019) se propone utilizar redes neuronales profundas (DNN) en su forma básica como base para la fusión de vistas. Posteriormente, se introduce el concepto de **regularizadores por correlación** (*Correlation Regularizers*, CorrReg), los cuales se aplican en las capas de fusión con el propósito de regularizar los parámetros y, de este modo, optimizar la validación del conjunto de datos. Continuando con la línea de la clasificación de imágenes utilizando redes neuronales, en el siguiente artículo titulado "*A Deep Multiview Active Learning for Large-Scale Image Classification*" (Tuozhong, Wenfeng, & Yuhong, 2020), se presenta el "Deep multiview active learning (DMAL) framework" el cual busca mejorar el aprendizaje activo de múltiples vistas (Multiview active learning (MAL)) esto realizando pruebas con variaciones de cantidad y tamaño de las imágenes, estas

acceden a dos redes neuronales convolucionales generando dos capas que se unen usando multiview learning para posteriormente tomar anotaciones de una persona y hacer retroalimentación activa a las redes neuronales y así complementar la información para su mejora en la clasificación.

Continuando con la línea de la clasificación de imágenes utilizando redes neuronales, en el siguiente artículo titulado "*A Deep Multiview Active Learning for Large-Scale Image Classification*" (Tuozhong, Wenfeng, & Yuhong, 2020), se presenta el "Deep multiview active learning (DMAL) framework" el cual busca mejorar el aprendizaje activo de múltiples vistas (Multiview active learning (MAL)) esto realizando pruebas con variaciones de cantidad y tamaño de las imágenes. Estas acceden a dos redes neuronales convolucionales generando dos capas que se unen usando multiview learning para posteriormente tomar anotaciones de una persona y hacer retroalimentación activa a las redes neuronales y así complementar la información para su mejora en la clasificación.

En la ilustración 5 tomada del artículo "*A Deep Multiview Active Learning for Large-Scale Image Classification*" (Tuozhong, Wenfeng, & Yuhong, 2020) se muestra cómo es el proceso en el cual las imágenes son analizadas en dos redes neuronales convolucionales, el resultado entra al modelo de multiview fusión el cual analiza las imágenes y define cuales puede entender y cuales no; posteriormente esta información se le presenta a una persona la cual hace una retroalimentación a las redes neuronales convolucionales, esto con la finalidad de mejorar los resultados que se obtienen en cada una de las redes neuronales convolucionales.

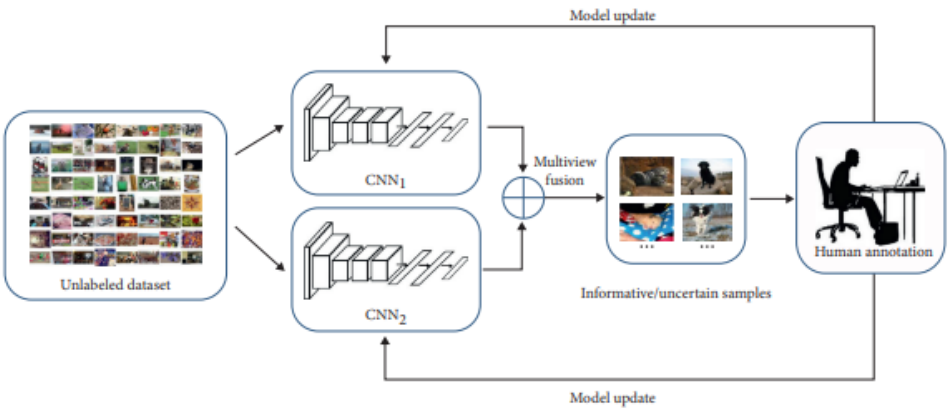

Ilustración 5: Muestra del proceso de multiview learning active. Fuente: "A Deep Multiview Active Learning for Large-Scale Image Classification" (Tuozhong, Wenfeng, & Yuhong, 2020)

El uso de redes neuronales convolucionales en combinación con el enfoque de MVL ha sido una línea de trabajo en crecimiento, con aplicaciones particularmente interesantes en campos especializados como la medicina. Un ejemplo relevante se presenta en el artículo "*Multi-View Convolutional Neural Networks for Mammographic Image Classification*" (SUN, WANG, HU, XU, & Cui, 2019), donde se aplica MVL para la clasificación de imágenes mamográficas. En este estudio, se emplean dos tipos de vistas médicas comunes: mediolateral oblicua (MLO) y cráneo-caudal (CC). A partir de estas vistas, se implementa una arquitectura denominada MVMDCNN, compuesta por dos subredes: una convolucional y otra específica para la vista oblicua medio lateral. Estas redes permiten extraer las características más representativas de las imágenes. Finalmente, se utilizan funciones de pérdida basadas en entropía cruzada para entrenar el modelo de manera profunda y precisa. Esta metodología se representa visualmente en la Ilustración 7, tomada del mismo artículo. (SUN, WANG, HU, XU, & Cui, 2019).

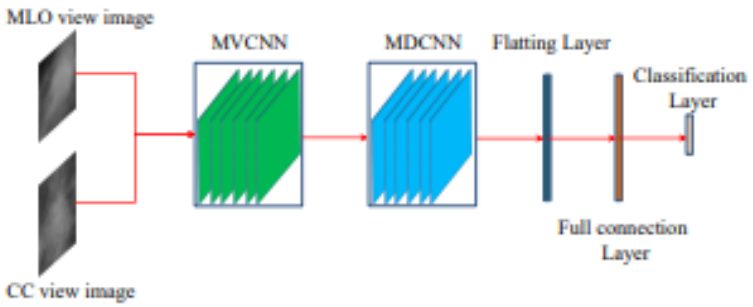

Ilustración 6: Muestra de la arquitectura MVMDCNN. Fuente: "Multi-View Convolutional Neural Networks for Mammographic Image Classification" (SUN, WANG, HU, XU, & Cui, 2019)

Es importante destacar que, detrás de estas estructuras de MVL existe una base teórica y algorítmica sustentada en diversos modelos de aprendizaje. Esta fundamentación permite explorar múltiples enfoques y *frameworks* específicos desarrollados en torno al MVL. Un ejemplo es el trabajo denominado: "*Multi-View Maximum Entropy Discrimination*" (Sun & Chao, 2013), en el que se propone un enfoque basado en la discriminación por máxima entropía (*Maximum Entropy Discrimination*, MED) implementado mediante *Support Vector Machines* (SVMs). El objetivo de este modelo es garantizar un margen de clasificación consistente entre las distintas vistas. Para validar su propuesta, los autores realizaron tres experimentos con distintos conjuntos de datos, demostrando que el modelo denominado MVMED ofrece mejoras significativas en comparación con otros métodos existentes. Una idea que se trabajó con MVL es el manejo de diccionarios de aprendizaje, en donde se utilizan estos dichos diccionarios para definir y encontrar características y así poder resolver problemas discriminando las imágenes como tal, esta técnica al ser usada tiene algunos aspectos a mejorar como el rendimiento el manejo del ruido en múltiples escenarios etc, es aquí donde se centra el artículo titulado: *"Multi-view low-rank dictionary learning for image classification"* (Wu, y otros, 2015; Wu, y otros, 2015), ya que realizan un trabajo con MVL supervisado que aprende sub diccionarios y aplica una fórmula de optimización para denotar la data presente en el mapa de la imagen y posteriormente correlacionarla discriminando posibles errores pero a un nivel bajo ya que regulariza algunos términos y haciendo más precisa la correlación. Una idea que se trabajó con MVL es el manejo de

diccionarios de aprendizaje, en donde se utilizan estos dichos diccionarios para definir y encontrar características y así poder resolver problemas discriminando las imágenes como tal, esta técnica al ser usada tiene algunos aspectos a mejorar como el rendimiento el manejo del ruido en múltiples escenarios etc, es aquí donde se centra el artículo titulado: *"Multi-view low-rank dictionary learning for image classification"* (Wu, y otros, 2015), ya que realizan un trabajo con MVL supervisado que aprende sub diccionarios y aplica una fórmula de optimización para denotar la data presente en el mapa de la imagen y posteriormente correlacionarla discriminando posibles errores pero a un nivel bajo ya que regulariza algunos términos y haciendo más precisa la correlación.

Continuando con las aplicaciones del MVL, se encuentra un trabajo enfocado al análisis de imágenes geográficas llamado *"Forest Fire Recognition Based on Feature Extraction from Multi-View Images"* (Wu, y otros, 2021), donde se hace el reconocimiento de incendios extrayendo características como los componentes que se encuentran en la imagen (humo y arboles). Se extraen características específicas, como los tonos rojos y amarillos asociados al fuego, a partir de un proceso dinámico de análisis de imágenes. Posteriormente, estas imágenes son procesadas mediante redes neuronales gráficas y convolucionales, lo que permite depurar aquellas que efectivamente contienen indicios de incendio. A partir de este filtrado, se estructuran dos capas fundamentales en el modelo: una destinada a la segmentación y otra a la clasificación del contenido visual.

MVL se ha utilizado para un trabajo de grado doctoral titulado "*Multi-View Imaging: Depth Estimation And Enhancement*" (Oliver, 2019) en el cual se aborda el manejo de múltiples vistas para soportar el "Stereo Problem" el cual es un problema de correspondencia entre dos imágenes separadas por una distancia horizontal mínima. Para ello se utilizan las redes neuronales junto con Multi-View Super-Resolution para lograr una mejora de las imágenes al hacer las correspondencias entre cada uno de los objetos.

Como se ha mencionado previamente, el uso de Multiview Learning (MVL) no se limita únicamente a la clasificación de objetos, sino que también se ha expandido al

reconocimiento de elementos a través de múltiples imágenes. Esto permite identificar si un objeto o persona, aunque presente variaciones entre imágenes, sigue siendo el mismo con base en un conjunto de características compartidas. Un ejemplo de esta aplicación se encuentra en el trabajo titulado *"Multiple View Oriented Matching Algorithm for People Reidentification"* (García, Gardel, Bravo, & Lázaro, 2014), donde se evidencia cómo, a partir de diferentes vistas, es posible identificar personas, reconocer cambios en su dirección y actualizar la perspectiva visual de manera continua. La ilustración 7 presenta un esquema conceptual que resume la base del modelo propuesto para esta tarea.

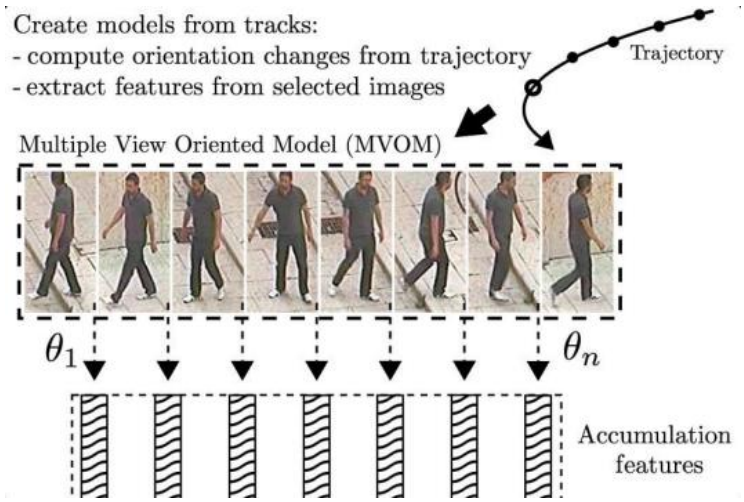

Ilustración 7: Muestra de Multiple View Oriented Matching Algorithm. Fuente: "Multiple View Oriented Matching Algorithm for People Reidentification" (García, Gardel, Bravo, & Lázaro, 2014)

MVL en su versión profunda "Deep multiview learning" se ha enfocado en mejorar tiempos y optimizar procesos. En el artículo *"Deep Multiview Learning for Hyperspectral Image Classification"* (Liu, y otros, 2021), se trabaja un método de aprendizaje profundo de múltiples vistas para abordar la clasificación de imágenes hiperespectrales (HSI), optimizando la cantidad de etiquetas que se usan para su respectiva clasificación, esto extrayendo características invariantes en cada vista para posteriormente entrar a una red no supervisada para su entrenamiento y posterior clasificación con Deep MVL.

El enfoque hiperespectral ha tomado gran relevancia en el contexto de MVL, ya que al involucrar múltiples dimensiones de información se busca mejorar tanto la precición como el rendimiento en la clasificación. En el artículo denominado *"Hyperspectral Image Classification via Spatial Window-Based Multiview Intact Feature Learning"* (Zhao, y otros, 2021), se propone el modelo "Spatial window-based multiview intact feature learning method" (SWMIFL) se basa en la extracción de características a partir de imágenes hiperespectrales y su posterior tratamiento mediante MVL con vistas intactas. Para complementar este proceso, se incorpora el método **Multiple Morphological Component Analysis (MMCA)**, que permite extraer de forma más precisa las características necesarias para el entrenamiento del sistema, el cual se implementa con máquinas de vectores de soporte (SVM).

Una de las aplicaciones destacadas de Deep MVL ha sido la segmentación multiclase en el análisis de ecografías, particularmente en el contexto del diagnóstico de cardiopatías congénitas (CHD). Se puede evidenciar en el trabajo *"Multiview and Multiclass Image Segmentation using Deep Learning in Fetal Echocardiography"* (Wong, Sinkovskaya, Abuhamad, & Mahmood, 2021), donde se propone una mejora del modelo V-Net mediante la incorporación de técnicas como abandono espacial (*spatial dropout*), normalización por grupos (*group normalization*) y supervisión profunda (*deep supervision*). Estas mejoras permiten una segmentación más precisa de las estructuras fetales al integrar información proveniente de múltiples vistas ecográficas, lo que contribuye a identificar anomalías que podrían pasar desapercibidas si se analiza una sola perspectiva.

En el artículo *"Multiview Similarity Learning for Robust Visual Clustering"* (Li, Chen, Chen, & Sun, 2020), se plantea un método robusto de aprendizaje de similitud de múltiples vistas (RMvSL), el cual busca implementar rangos bajos en las muestras de las imágenes para eliminar ruido y valores atípicos, de tal forma que puedan establecer un esquema de múltiples vistas que pueda trabajar bajo una función de optimización que realice la

reconstrucción de los datos y sirva para el aprendizaje a la hora de fusionar diferentes gráficos.

Otra posibilidad que se ha abierto con MVL, es la de realizar la concatenación de información que no se relaciona a simple vista. Un ejemplo de ello es el trabajo *"Reconstructing Perceived Images From Human Brain Activities With Bayesian Deep Multiview Learning"* (Du, Du, Huang, & He, 2018), en el cual se crea un módulo multivista enfocado a la reconstrucción de imágenes visuales, y que además, se realiza a partir de la actividad cerebral registrada en una resonancia magnética funcional que en el entrenamiento, logró relacionar estadísticamente las dos vistas (la resonancia y la imagen que es estimulo visual), de tal forma que al momento de obtener una resonancia se pueda obtener el estímulo visual que se aplicó mediante una red neuronal profunda. En la ilustración 8, se puede ver la idea general del desarrollo de este artículo, siendo la primera parte del entrenamiento cómo se relacionan las dos vistas y cómo en la prueba de predicción se deben comportar las actividades cerebrales para posteriormente reconstruir el estímulo esperado.

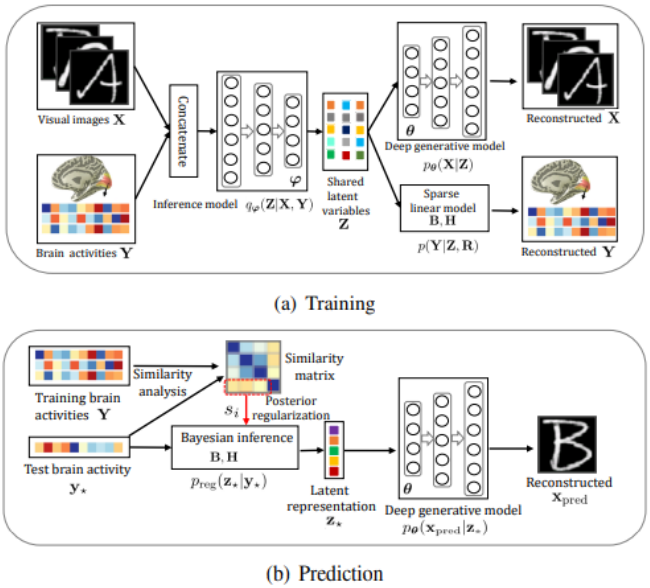

Ilustración 8: Muestra de Entrenamiento y predicción del modelo de decodificación Neuronal. Fuente: "Reconstructing Perceived Images From Human Brain Activities With Bayesian Deep Multiview Learning" (Du, Du, Huang, & He, 2018)

Una ventaja que se ha encontrado utilizando MVL para el manejo de imágenes es la capacidad de preservar las características de los datos al momento de ser extraídas de la fuente original. Esto es evidente en la investigación "Deep Multi-View Enhancement Hashing for Image Retrieval" (Yan, Gong, Wei, & Gao, 2020), en donde se crea una red neuronal profunda multivista, que se enfoca en el aprendizaje del hash (del vecino más cercano en un espacio de dato) para así lograr recuperar la información de una forma eficiente, y así estar evaluando activamente la relación entre las vistas ya que tienen en cuenta el ruido como parte de la entrada. En la Ilustración 10 se evidencia como es el proceso de tratamiento de la información en este trabajo, en donde dicho marco propuesto de hash de mejora de múltiples vistas profundas (D-MVE-Hash).

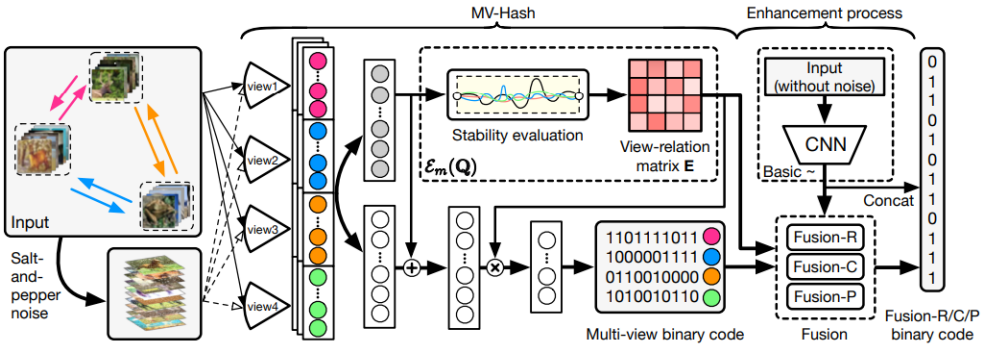

Ilustración 9: (D-MVE-Hash) framework. Fuente: "Deep Multi-View Enhancement Hashing for Image Retrieval" (Yan, Gong, Wei, & Gao, 2020)

Una situación recurrente en el manejo de imágenes es cuando las imágenes poseen alguna etiqueta previa, lo que ha motivado el desarrollado diversos trabajos enfocados a la comprensión de dichas etiquetas. Sin embargo, existe una tendencia a descuidar las interacciones intramodales que puede haber entre las mismas etiquetas. En contexto, el artículo *"Multimodal Transformer with Multi-View Visual Representation for Image Captioning"* (Yu, Li, Yu, & Huang, 2020), plantea un modelo Transformador multimodal (MT) el cual permite la comparación de las etiquetas que han sido asignadas a las imágenes preexistentes, y combina las interacciones intramodales e intermodales con un razonamiento multimodal complejo para generar etiquetas precisas y acordes al contexto

que se está presentando. Se puede observar en la Ilustración 10, donde para una imagen se usa un detector que asigna las etiquetas las cuales generan m vistas que posteriormente entraran a un codificador para así lograr asignar correctamente la etiqueta.

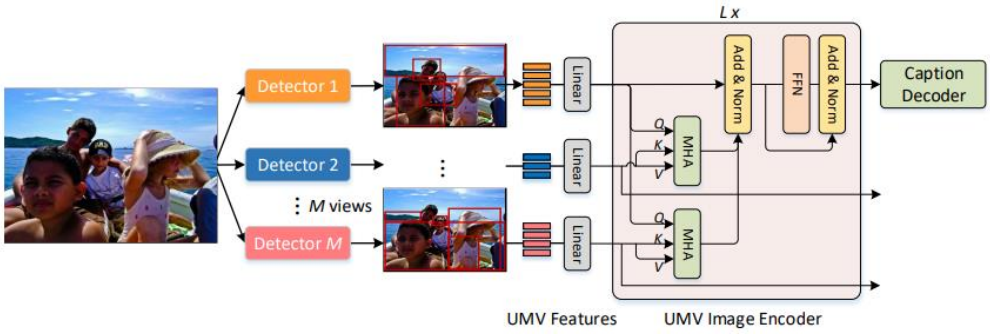

Ilustración 10: Muestra de la estructura del manejo de etiquetas Multimodal. Fuente: "Multimodal Transformer with Multi-View Visual Representation for Image Captioning" (Yu, Li, Yu, & Huang, 2020)

Profundizando en las aplicaciones médicas de MVL, el trabajo titulado *"Multi-view Learning and Deep Learning for Microscopic Neuroblastoma Pathology Image Diagnosis"* (Liu, Sun, & Yin, 2018)ofrece un enfoque innovador en el diagnóstico de neuroblastoma a partir de imágenes microscópicas. Se parte del objetivo de detectar patologías oncológicas mediante el análisis automatizado de imágenes, centrándose en el neuroblastoma, el tumor sólido extracraneal más frecuente en la infancia. El proceso diagnóstico requiere la clasificación de imágenes en diferentes subtipos: indiferenciado (UD), pobremente diferenciado (PD), diferenciador (D) y tejido normal no neoplásico (UN). Se comienza haciendo una extracción de las características para generar las múltiples vistas con algoritmos de bajo nivel como lo son LBP y DSIFT; posteriormente, se hace una función de clasificación basada en una función discriminante de las múltiples vistas en la parte de MVMED Classifier donde se hace la fusión de las clasificaciones respectivas y finalmente, se da un resultado final de la clasificación tal cual como se muestra en la Ilustración 12.

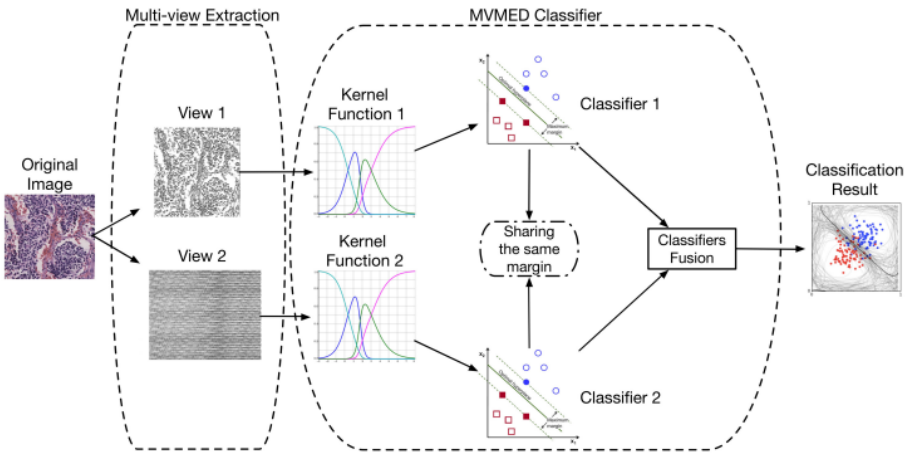

Ilustración 11: Muestra de la estructura de la arquitectura del MVMED. Fuente: "Multi-view Learning and Deep Learning for Microscopic Neuroblastoma Pathology Image Diagnosis" (Liu, Sun, & Yin, 2018)

Para concluir la revisión de artículos relacionados con MVL aplicado a imágenes y sus temáticas asociadas, se revisó el estudio *"Embedding Regularizer Learning for Multi-View Semi-Supervised Classification"* (Huang, Wang, Zheng, Zhao, & Lin, 2021), en el cual tiene como objetivo establecer un marco metodológico que permita incorporar regularización de información con el fin de integrar información complementaria entre las diferentes vistas. El enfoque se basa en una matriz de etiquetas para facilitar la clasificación multivista semisupervisada, a través de un esquema denominado *Embedding Regularizer Learning for Multi-View Semi-Supervised Classification* (ERL-MVSC). La propuesta incluye una estrategia de optimización conjunta, que combina la regularización de la incrustación de información y la predicción de etiquetas, utilizando un método de descenso por coordenadas para resolver el problema de forma eficiente.

2.2.3 Marco teórico enfocado al conflicto armado:

Debido a que la selección del caso de estudio fue enfocado a la temática del conflicto armado se hace un énfasis en tener un marco teórico acerca del conflicto armado en donde se resalten temas que den contexto para la definición de las imágenes como violentas o no violentas dentro del marco del conflicto armado en Colombia.

Para ello se abordan diversos textos que van desde los fundamentos generales en los que se basa el conflicto armado, pasando por los actores armado (grupos al margen de la ley y estado), su impacto en la sociedad y las victimas, hasta factores residuales como son el hecho de la memoria (principal eje del aspecto de manejo de imágenes relacionadas con el conflicto) y la reconciliación como ultima consecuencia del conflicto armado.

El conflicto armado en Colombia es un tema que en si contiene muchas vertientes y a su vez hablar de ello genera diversos puntos de vista, por tal motivo es necesario hacer un repaso objetivo sobre el contexto general sobre el cual se desarrolla este marco teórico; el conflicto armado en Colombia, que lo enmarcaremos desde la mitad del siglo XX constituye un escenario complejo en el que confluyen actores armados, intereses políticos y factores económicos que han transformado de manera profunda a la sociedad. Este proceso no puede entenderse únicamente como una confrontación militar, sino como una guerra que ha fragmentado territorios, erosionado instituciones y situado a la población civil en el centro de la violencia. En este marco, la investigación académica ha desarrollado múltiples perspectivas que permiten aproximarse a sus dinámicas: desde las interpretaciones teóricas del conflicto como una disputa por soberanías y órdenes locales como lo plantean autores como (Pécaut, 2001) con orígenes económicos como lo plantea (Rettberg & Prieto, 2018), hasta los estudios sobre la relación entre los grupos armados al margen de la ley, política y Estado (Gutiérrez Sanín, 2019)

Al mismo tiempo, se ha puesto un énfasis creciente en el impacto social y las experiencias de las víctimas, ampliamente documentadas por el Centro Nacional de Memoria Histórica

y por autores como Sánchez y Rettberg, quienes destacan la importancia de la memoria y la reconciliación como dimensiones centrales para la construcción de paz. (Centro de memoria historica de Colombia, 2025)

Otros trabajos, como los de Karl o Arjona, han subrayado la relevancia de factores estructurales —económicos y territoriales— que prolongaron y complejizan el conflicto. Esta diversidad de aproximaciones ofrece no solo un panorama rico para el análisis académico, sino también un marco de referencia fundamental para comprender los contextos en los que emergen las imágenes del conflicto.

El acervo visual resultante como lo son las fotografías, registros periodísticos y documentales gráficos constituye una fuente de memoria, pero también plantea desafíos para su análisis sistemático. En la actualidad, el campo del aprendizaje automático ha buscado responder a estos retos mediante técnicas de clasificación de imágenes violentas, que permiten procesar grandes volúmenes de datos visuales. Sin embargo, gran parte de estas aproximaciones se han concentrado únicamente en el plano visual, dejando de lado el trasfondo histórico y social que dota de sentido a dichas imágenes.

En este punto, el enfoque de *multiview learning* representa una oportunidad para avanzar hacia modelos que integren distintas dimensiones de la información: no solo el contenido visual, sino también metadatos, descripciones textuales y referencias contextuales derivadas de la literatura académica. De este modo, el diálogo entre estudios sociales del conflicto y técnicas de visión por computador abre un espacio interdisciplinar en el que la tecnología no se limita a clasificar imágenes, sino que contribuye a interpretarlas en clave histórica y social, fortaleciendo tanto el análisis científico como la construcción de memoria.

A partir de esta premisa se realiza una revisión bibliográfica que permite observar no solo la dinámica del conflicto en el país, sino también para analizar los contextos en donde se las secuencias violentas adquieren su sentido.

En el trabajo *Guerra de la sociedad* (Pécaut, 2001) plantea un panorama sobre la violencia en Colombia, el cual no entiende el conflicto como un enfrentamiento entre grupos armados sino una guerra entre diferentes actores grupos al margen de la ley, las mismas poblaciones e incluso entes del estado, que como finalidad atentan contra el bienestar y el tejido social que la sociedad puede llegar a construir. En su análisis, menciona diferentes elementos fundamentales que califican y definen las intenciones detrás de la perpetuación del conflicto interno en el país a través de las décadas, en donde se resaltan aspectos como la violencia como herramienta de control territorial y económico en los poblados locales que están por fuera del control estatal mostrando la incapacidad de las instituciones estatales para generar protección y una estructura social consolidada.

Avanzando en el análisis encontramos el trabajo *Guerras, memoria e historia* (Sánchez Gómez & Perea Restrepo, 2006) , donde el autor aborda el conflicto en el país, no desde los enfrentamientos entre grupos armados, sino desde la memoria histórica. En su trabajo, el autor desarrolla tres ideas fundamentales, planteando es que los procesos de reconciliación que se presentan tras los momentos de conflicto no son actos *individuales* o de experiencia individual, ya que estos procesos requieren de la participación de todas las personas como pueblo y no como individuo, para lograr la construcción común de la paz. Después, el autor menciona la historia del conflicto como una dicotomía, donde los discursos -de distintos pares- funcionan como un sistema de deslegitimación para el grupo contrario. Por lo cual, se debe entender que no existe una sola memoria, sino que existe la versión oficial y la comunitaria, donde una busca demeritar la otra.

Por su parte en el trabajo *"Parapolítica: la ruta de la expansión paramilitar y los acuerdos políticos*" (Romero, 2007) realiza un recorrido donde habla sobre la forma en que operan los grupos armados. Especialmente, menciona cómo estas estructuras no solo trabajan efectuando atentados y llevando a cabo combates con la fuerza pública, sino que también fungen como alianzas con diferentes sectores políticos y económicos de Colombia. Así, su análisis permite entender cómo las fuerzas de los grupos armados al margen de la ley

funcionaron como un proyecto político-militar que facilitó la expansión territorial de los grupos armados.

En una línea común, en el trabajo *Clientelistic Warfare: Paramilitaries and the State in Colombia* (Gutiérrez Sanín, 2019) realiza un análisis del conflicto haciendo énfasis en los vínculos entre diferentes sectores del estado y los distintos grupos armados que actúan en el país. Especialmente, el autor profundiza en la idea de que los grupos militares no se deben entender como grupos armados al margen de la ley, sino como una forma diferente de gobierno, con una forma diferente de mandato que puede incluir distintas formas de violencia o acuerdos económicos y políticos.

Desde un ámbito académico, como se plantea en *"Los estudios sobre la violencia: balance y perspectivas"* (Sánchez, 2007) se aborda el conflicto armado realizando un recorrido por el estado del arte de la guerra en Colombia. En su trabajo, realizan un recorrido histórico donde se logra evidenciar cómo el conflicto armado se entiende inicialmente como un problema rural únicamente; y cómo durante muchos años, la selva se consideró como el único escenario donde se perpetúan las situaciones de violencia, así como los grupos armados como únicos actores. Sin embargo, con el paso de los años, los autores mencionan que los escenarios y los actores se han desplazado a diferentes escenarios urbanos como pueblos y pequeñas ciudades –y en el gobierno Incluso–, por lo que es necesario pensar el conflicto como una situación multidisciplinar que aborda aspectos sociales, económicos, políticos y simbólicos.

El análisis de la literatura revisada nos muestra que la dinámica del conflicto armado en Colombia no es unidimensional, es decir, no debe pensarse únicamente desde la perspectiva de las armas y los enfrentamientos, sino que debe comprenderse todo un grupo de factores tanto históricos, económicos y políticos que son los que han desatado diferentes episodios de violencia que ha vivido el país. De igual forma, las investigaciones nos muestran que el concepto de *conflicto* o *violencia* no hace referencia exclusivamente a grupos armados y a zonas rurales, sino que también se manifiesta a través de disputas territoriales,

el olvido histórico o la omisión gubernamental, y que también se puede encontrar en los epicentros del país. En conclusión, entender el contexto del conflicto colombiano desde sus inicios, actores, impacto y consecuencias; ayuda a esta investigación a brindar un contexto más adecuado sobre el cual se pueden trabajar las imágenes relacionadas con el conflicto armado en Colombia, siendo esto necesario para la fase de búsqueda, selección y preclasificación de las imágenes para el apartado de validación y pruebas de modelo.

Para ello es importante ver que trabajos se han abordado el manejo de imágenes y el conflicto armado, tanto por su contexto como por su forma de clasificación, para ello se adentró en temas como la importancia de las imágenes como un medio de entender el conflicto, como se muestra en el artículo *"La imagen en el contexto de la violencia en Colombia: un acercamiento a distintas perspectivas"* (Castaño Giraldo, Avella Estrada, Arango Murcia, & Sánchez, 2016) y como las imágenes se convierten en herramientas para representar, recordar, resistir y resignificar la violencia en el país.

Cabe resaltar la importancia de las imágenes que la sociedad colombiana consume y uno de sus principales actores son los medios de comunicación sobre todo impresos como periódicos y revistas, sobre los cuales se revisó el articulo *"Memorias visuales del conflicto armado en Colombia (2008–2016): un análisis sobre los diarios El Tiempo y El Espectador"*.

2.3 Marco Conceptual

Con la idea de establecer las nociones relacionados con las temáticas a trabajar se establecieron estos conceptos los cuales son fundamentales para el entendimiento del marco teórico

- **Multiview learning** (**MVL**) es una rama del "machine learning" que tiene un especial enfoque por mejorar el rendimiento y sus resultados al trabajar múltiples vistas, "El aprendizaje de múltiples vistas también se conoce como fusión de datos o integración de datos de múltiples conjuntos de características." (Zhao, Xijiong, Xin, & Shiliang, 2017) y que a su vez es un concepto relativamente nuevo que se ha ido explorando durante la última década con diferentes enfoques, algoritmos y estrategias para el entrenamiento que dependen de los problemas a resolver.

  Otra definición en la literatura académica la suele referenciarla como "El aprendizaje de múltiples vistas es una dirección emergente en el aprendizaje automático que considera el aprendizaje con múltiples vistas para mejorar el rendimiento de la generalización. El aprendizaje multivista también se conoce como fusión de datos o integración de datos de múltiples conjuntos de características." (Jing Zhao, 2017)
- **Redes convolucionales:** Cuando se trabaja con imágenes en el contexto del aprendizaje automático, es fundamental contar con modelos capaces de interpretar no solo los valores de los píxeles, sino también la estructura y distribución espacial de los elementos en la imagen. En este sentido, las redes neuronales convolucionales (CNN) surgen como una solución particularmente eficaz. A diferencia de las redes tradicionales, que suelen requerir aplanar las imágenes en vectores unidimensionales (perdiendo así su organización espacial), las CNN aplican una operación matemática llamada *convolución*, la cual permite identificar patrones locales en la imagen sin sacrificar su forma original. Tal como se explica en el trabajo denominado *"Redes neuronales convolucionales aplicadas a la identificación y medición automatizadas"* (Francés Falip, 2020), este tipo de redes no solo preserva

la información espacial, sino que también optimiza el uso de parámetros mediante la reutilización de pesos, lo que mejora su rendimiento computacional. Además, las CNN están preparadas para trabajar con imágenes en color y múltiples canales, lo que amplía su aplicabilidad en contextos reales. Esta capacidad de detectar características relevantes, como bordes o texturas, directamente desde la imagen, las convierte en herramientas especialmente útiles en tareas de clasificación visual, como las que se abordan en esta investigación.

- **Análisis de textos:** el análisis de texto ha demostrado ser un complemento fundamental al análisis visual, especialmente en tareas donde la apariencia externa del documento no es suficiente para una clasificación precisa. Este tipo de análisis implica la extracción del contenido textual de imágenes mediante técnicas de reconocimiento óptico de caracteres (OCR), seguido por una representación semántica del texto (frecuentemente a través de *word embeddings*) que permite al modelo interpretar su significado. Un trabajo interesante es *"Multimodal deep networks for text and image-based document classification"* (Audebert, Herold, Slimani, & Vidal, 2019), el cual muestra los enfoques que integran tanto las características visuales como las textuales en una arquitectura multimodal logran mejorar significativamente el rendimiento respecto a modelos que utilizan solo una de estas fuentes de información. En su estudio, los autores muestran que el uso combinado de imágenes y texto extraído permite superar en precisión a los modelos unicanal hasta en un 3 %, incluso cuando el texto proviene de OCRs con ruido. Esto subraya la relevancia del análisis de texto como una vía efectiva para captar matices semánticos que no están presentes en la estructura visual del documento, y refuerza su utilidad en entornos de clasificación donde se requiere un reconocimiento fino de clases.

  Es importante considerar los desafíos que surgen cuando el análisis textual debe realizarse sobre imágenes capturadas en entornos no estructurados o no controlados, como ocurre con los documentos escaneados por usuarios o con

fotografías tomadas en escenas cotidianas. Este tipo de texto, conocido como *texto en escena*, introduce variabilidad significativa en términos de orientación, tipografía, iluminación y legibilidad. En este sentido, el trabajo "*COCO-Text: Dataset and Benchmark for Text Detection and Recognition in Natural Images*" (Veit, Matera, Neumann, Matas, & Belongie, 2016), realiza la creación del conjunto de datos COCO-Text representan un avance relevante, al proporcionar una base diversa y extensa para entrenar y evaluar algoritmos de detección y reconocimiento de texto en imágenes naturales. Los autores destacan que, si bien los sistemas actuales logran buenos niveles de precisión en el reconocimiento de texto legible, aún presentan debilidades considerables en *recall*, especialmente frente a textos manuscritos o parcialmente visibles. Estas limitaciones evidencian que el análisis textual dentro de tareas de clasificación de imágenes no puede abordarse como un proceso aislado, sino como parte de una estrategia integral que combine herramientas de visión por computador, técnicas de OCR y modelos de representación semántica robustos. Esta integración no solo mejora la precisión en escenarios reales, sino que también amplía el alcance de las aplicaciones basadas en análisis multimodal.

- **Redes neuronales:** Las redes neuronales artificiales (ANNs) son modelos computacionales inspirados en el funcionamiento del sistema nervioso biológico, particularmente en la forma en que las neuronas del cerebro humano procesan la información. Estas redes están compuestas por nodos interconectados (llamados comúnmente "neuronas") que colaboran en forma distribuida para aprender patrones a partir de los datos de entrada, ajustando sus conexiones internas con el fin de optimizar los resultados. Un trabajo que presenta las redes neuronales es el trabajo *"An Introduction to Convolutional Neural Networks"* (O'Shea & Nash, 2015),en donde el aprendizaje en una red neuronal ocurre a través de la retroalimentación que cada capa proporciona a la anterior, evaluando cómo pequeñas modificaciones en sus parámetros impactan en la salida final. Este proceso iterativo permite que las redes aprendan representaciones complejas de los

datos, siendo especialmente útiles en tareas de clasificación, regresión y reconocimiento de patrones. Cuando se utilizan múltiples capas ocultas, se habla de "aprendizaje profundo", una estrategia que ha potenciado significativamente el rendimiento de los sistemas de inteligencia artificial en los últimos años. Además de las redes neuronales tradicionales, las redes neuronales convolucionales (CNN) han demostrado ser particularmente eficaces en tareas de clasificación de imágenes. A diferencia de los modelos que requieren una extracción manual de características como texturas o formas, las CNN tienen la capacidad de aprender automáticamente estas representaciones directamente a partir de los datos de píxeles sin procesar. Este enfoque permite que el modelo identifique patrones jerárquicos en las imágenes, extrayendo progresivamente características de mayor nivel que facilitan la identificación de objetos y estructuras complejas. Según lo expuesto por Google Developers, una CNN procesa una imagen de entrada mediante una serie de capas que aplican filtros convolucionales, generando mapas de características que capturan distintos aspectos del contenido visual. Este mecanismo no solo mejora la precisión en la clasificación, sino que también reduce la necesidad de preprocesamiento manual, haciendo que las CNN sean una herramienta poderosa y eficiente en el campo del aprendizaje automático aplicado a la visión por computadora.

- **Reconocimiento de objetos:** Dentro del ámbito de la visión artificial, el reconocimiento de objetos se refiere a la capacidad de un sistema para no solo identificar qué objetos están presentes en una imagen, sino también para determinar con precisión su posición dentro de la misma. Esta tarea implica detectar instancias específicas de interés y asociarlas a una categoría definida, todo esto a través de una representación visual que delimita su ubicación, generalmente mediante cuadros delimitadores. A nivel técnico, este proceso comienza con una preparación minuciosa del conjunto de datos, donde se recopilan y etiquetan imágenes de manera estandarizada para garantizar que el modelo aprenda de forma

efectiva. Posteriormente, se lleva a cabo el entrenamiento del modelo empleando algoritmos de aprendizaje supervisado, con el fin de que la red pueda aprender a reconocer patrones visuales consistentes asociados a cada clase de objeto. Finalmente, el modelo entrenado se aplica a nuevas imágenes para realizar la detección, generando predicciones sobre la presencia, ubicación y categoría de cada objeto detectado. Tal como se plantea en el artículo *"Sistema de reconocimiento de patrones en imágenes satelitales para la detección de objetos"* (Fernández Marcos, 2023), el uso de redes neuronales convolucionales en este tipo de tareas ha demostrado ser especialmente eficaz debido a su capacidad para extraer características visuales relevantes incluso en condiciones complejas, como cambios en la escala o en la orientación de los objetos.

- **Machine learning:** El aprendizaje automático (*Machine Learning*) es una rama central de la inteligencia artificial que busca dotar a las computadoras de la capacidad de aprender de los datos y mejorar su rendimiento sin necesidad de una programación explícita para cada tarea. A través de este enfoque, los sistemas son capaces de adquirir nuevos conocimientos, reconocer patrones existentes y perfeccionar su comportamiento con el tiempo. Esto se presenta en el artículo *"Machine learning and applications: A review study. International Journal of Multidisciplinary Research and Analysis"* (Atiyah1 & Hassan2, 2024), el cual plpantea que el objetivo fundamental del aprendizaje automático es simular el proceso de aprendizaje humano mediante algoritmos que transforman información del entorno en conocimiento útil y accionable. Este proceso se apoya en un modelo estructurado que comprende la recolección de datos desde el entorno, su transformación en representaciones internas almacenadas en repositorios, y su uso para realizar tareas específicas, retroalimentando el sistema para futuras mejoras. La versatilidad del aprendizaje automático y su capacidad para adaptarse a diferentes dominios explican su creciente adopción en sectores como las telecomunicaciones, el análisis financiero, la medicina y la ingeniería, entre otros.

Al ser un concepto que tiene una amplia definición, una de ellas es "disciplina científica del ámbito de la Inteligencia Artificial que crea sistemas que aprenden automáticamente. Aprender en este contexto quiere decir identificar patrones complejos en millones de datos." (Gonzáles, 2014).

- **Deep Learning:** a nivel conceptual se pude definir de la siguiente manera "Como un algoritmo automático estructurado o jerárquico que emula el aprendizaje humano con el fin de obtener ciertos conocimientos. Destaca porque no requiere de reglas programadas previamente, sino que el propio sistema es capaz de "aprender" por sí mismo para efectuar una tarea a través de una fase previa de entrenamiento." (smartpanel, 2019).
- **Visualización de datos:** Al ser un concepto que se relaciona con los sistemas de visualización, se puede dar un contexto como "La presentación de datos en un formato pictórico o gráfico… describe la presentación de la información abstracta en forma gráfica. Permite detectar patrones, tendencias y correlaciones que de otro modo podrían pasar desapercibidos en los informes, tablas o hojas de cálculo tradicionales." (Tecnologías información, 2018)
- **Sistemas de clasificación supervisados y no supervisados:** El uso de técnicas de clasificación en inteligencia artificial ha cobrado gran relevancia en diversos campos de investigación, particularmente en aquellos donde resulta fundamental analizar grandes volúmenes de datos para facilitar la toma de decisiones automatizada. Dentro de este contexto, destacan dos enfoques ampliamente empleados: la clasificación supervisada y la no supervisada. A lo largo de este apartado se examina, desde una perspectiva técnica con orientación práctica, cómo se implementan estos métodos en distintas investigaciones científicas pertenecientes a diferentes áreas del conocimiento.

  La clasificación supervisada se distingue por requerir conjuntos de datos previamente etiquetados; es decir, cada muestra debe estar asignada a una categoría ya establecida. Esta información permite que los algoritmos identifiquen

patrones y relaciones entre las variables, a fin de predecir la categoría de nuevas observaciones. En cambio, la clasificación no supervisada opera sobre datos no etiquetados, buscando descubrir estructuras ocultas o agrupamientos naturales sin contar con información previa sobre las clases existentes.

En otro ámbito se presenta el trabajo *"Clasificación de hojas de té al ingreso del proceso de secado mediante redes neuronales con datos supervisados y no supervisados"* (Paula, Eckert, & Guismín, 2018), en donde se aplicaron ambas técnicas de clasificación en el proceso de selección de hojas de té al ingreso a los secaderos. En este caso, se utilizó clasificación supervisada mediante redes neuronales artificiales entrenadas con imágenes previamente etiquetadas por expertos. Al mismo tiempo, se implementó el algoritmo K-means para llevar a cabo un agrupamiento no supervisado. Esta integración permitió no solo una clasificación precisa, sino también el reconocimiento de patrones en los datos sin necesidad de intervención humana directa, facilitando la adaptación del sistema a nuevas muestras sin requerir etiquetado manual.

En el ámbito de la salud, se revisó el articulo "Clasificación no supervisada de imágenes médicas y minería de datos. Algoritmo S3 vs K-medias" (Sánchez Álvarez, 2021), en el cual se desarrolló una investigación centrada en la clasificación de imágenes médicas utilizando exclusivamente métodos no supervisados. Su trabajo comparó el algoritmo K-means con un nuevo modelo denominado S3 (de tres pasos), diseñado para manejar eficientemente datos de gran dimensión mediante programación paralela. Ambos algoritmos se aplicaron a la segmentación de imágenes, y el algoritmo S3 demostró ventajas sustanciales al no depender de la elección inicial de centroides, aumentando así la robustez del sistema frente a la variabilidad de los datos.

Por su parte, los autores del articulo *"Classification of lidar measurements using supervised and unsupervised machine learning methods"* (Farhani, Sica, & Daley, 2020), aplicaron tanto técnicas supervisadas como no supervisadas para clasificar

datos provenientes de mediciones LIDAR. En el enfoque supervisado, entrenaron modelos como Random Forest, SVM y Gradient Boosting para categorizar perfiles en “buenos”, “malos” o “nublados”, optimizando así la detección de perfiles válidos. En paralelo, utilizaron el algoritmo t-SNE como técnica no supervisada para agrupar perfiles similares y detectar anomalías, como la presencia de aerosoles estratosféricos derivados de incendios forestales. Esta estrategia combinada permitió no solo una clasificación eficiente, sino también una exploración detallada de patrones especiales dentro de los datos.

Finalmente, en el estudio *“Predicción del rendimiento académico en la UNADECA por medio de sistemas de clasificación”* (Castillo Aráuz & Martínez, 2023), centrado en la predicción del rendimiento académico en la Universidad Adventista de Centroamérica, se implementaron exclusivamente técnicas de clasificación supervisada. Utilizando algoritmos como Random Forest y Gradient Boosted Trees, los autores lograron predecir con más del 90 % de precisión si un estudiante aprobaría o reprobaría un curso. Esta investigación aplicó minería de datos educativos y aprendizaje automático sobre datos históricos etiquetados, demostrando el potencial de estos modelos para apoyar la toma de decisiones académicas y diseñar políticas educativas personalizadas.

En síntesis, los modelos de clasificación supervisada resultan particularmente útiles en contextos donde se dispone de datos etiquetados y se busca una alta precisión predictiva. Por otro lado, los métodos no supervisados son más adecuados para explorar conjuntos de datos amplios y sin estructura previa. La aplicación combinada de ambos enfoques, como se evidencia en varias de las investigaciones revisadas, permite abordar los desafíos contemporáneos en ciencia de datos con mayor flexibilidad y efectividad.

- **Esquemas de evaluación y valoración:** la evaluación del software constituye un proceso fundamental para determinar si un producto, herramienta o modelo computacional cumple con los objetivos para los cuales fue diseñado, tanto desde

una perspectiva técnica como funcional. En este contexto, los esquemas de valoración y evaluación de modelos de software son conjuntos estructurados de criterios, métodos y métricas que permiten analizar de forma sistemática y objetiva la calidad y el rendimiento de un sistema. Su aplicación abarca desde la etapa de desarrollo hasta la implementación y mantenimiento, con el propósito de garantizar productos robustos, confiables y alineados con las necesidades de los usuarios y del entorno organizacional.

En términos generales, un esquema de **valoración** está orientado a establecer juicios cualitativos o cuantitativos sobre un modelo o herramienta, considerando características como utilidad, adecuación al contexto o cumplimiento de estándares. Por otro lado, un esquema de **evaluación** implica un proceso más sistemático y medible, en el que se aplican modelos normativos, métricas específicas y técnicas de verificación para comparar un producto contra criterios establecidos o contra otros productos similares. Ambas nociones, aunque complementarias, cumplen funciones distintas: la valoración tiende a ser más interpretativa, mientras que la evaluación exige evidencia y resultados medibles.

Un ejemplo ampliamente reconocido de esquema de evaluación es el modelo ISO/IEC 9126, que estructura la calidad del software en seis dimensiones: funcionalidad, fiabilidad, usabilidad, eficiencia, mantenibilidad y portabilidad; esto se encuentra en el documento *"Guía técnica para la evaluación de software."* (Largo Garcia & Marin Mazo, 2009). Este modelo no solo establece qué evaluar, sino también cómo hacerlo, mediante indicadores y métricas adaptables al tipo de producto. A partir de esta base, han surgido otros modelos que complementan o expanden estos criterios.

Una evolución importante de estos enfoques se encuentra en el marco propuesto en el trabajo *"Classification and Evaluation Framework of Automated testing tools*

for agile software: Technical Review" (Mosleh, Al-Khulaidi, Gumaei, Alsabry, & Musleh, 2024), enfocado en la clasificación y evaluación de herramientas de prueba automatizada dentro de metodologías ágiles. En este caso, el esquema de evaluación contempla criterios específicos para entornos cambiantes: integración con pipelines de CI/CD, cobertura de pruebas, facilidad de scripting y compatibilidad con herramientas colaborativas. Se logra observar cómo los esquemas de evaluación deben ser flexibles y adaptables a las prácticas actuales de desarrollo.

Por otro lado, en el trabajo *"Evaluation framework for systems models. CPT: Pharmacometrics & Systems Pharmacology"* (Braakman, Pathmanthan, & Moore, 2022), se aporta un marco para evaluar modelos matemáticos complejos, en particular, en farmacología computacional. Este esquema enfatiza el uso de técnicas como el análisis de sensibilidad, la cuantificación de incertidumbre y la validación estadística para asegurar que los modelos sean confiables incluso cuando no se dispone de datos experimentales para contrastar sus predicciones. La evaluación en este contexto no se limita a un resultado binario de aprobación o rechazo, sino que implica un proceso profundo de interpretación y documentación.

Más recientemente, frente a los modelos de lenguaje grande (LLM), se han requerido nuevos enfoques de evaluación. Según *"Evaluating LLMs: complex scorers and evaluation Frameworks"* (Symflower, s.f.),los esquemas tradicionales resultan insuficientes ante modelos generativos cuyo comportamiento no siempre puede medirse mediante una sola respuesta correcta. Por ello, proponen una clasificación entre métricas supervisadas (usadas cuando hay una verdad de referencia) y no supervisadas la cual es adecuadas cuando el modelo produce múltiples salidas válidas. También incorporan evaluadores basados en otros modelos (LLM-as-a-judge), y métricas semánticas como BERTScore o SelfCheckGPT, que valoran la coherencia y relevancia del texto generado. Estas herramientas representan una

forma moderna y automatizada de aplicar esquemas de valoración y evaluación en escenarios complejos.

Además, la herramienta DevQualityEval, también desarrollada por Symflower, introduce un enfoque pragmático para evaluar modelos de lenguaje aplicados al desarrollo de software. Su esquema considera desde el costo por token hasta la calidad del código generado, incluyendo aspectos como legibilidad, cumplimiento de buenas prácticas y pertinencia de las funciones propuestas. Este tipo de evaluación híbrida refleja cómo los esquemas actuales deben conjugar aspectos técnicos, económicos y funcionales en un solo marco.

En conclusión, los esquemas de valoración y evaluación de modelos de software son herramientas indispensables para asegurar la calidad y pertinencia de los productos tecnológicos que desarrollamos o implementamos. Desde modelos normativos hasta frameworks adaptativos, cada uno aporta elementos clave para el análisis, la comparación y la mejora continua de herramientas y sistemas. Entender sus diferencias y aplicaciones nos permite no solo elegir mejores soluciones, sino también generar procesos de desarrollo más responsables, eficientes y centrados en el usuario final.

- **Métricas de evaluación:** En el contexto del aprendizaje automático, la evaluación del rendimiento de los modelos de clasificación es una etapa fundamental para determinar su eficacia ante datos nuevos. Esta evaluación permite identificar qué tan bien un modelo es capaz de predecir correctamente las clases de las instancias, lo cual es especialmente importante en tareas sensibles como el análisis de texto, la detección de fraudes o el diagnóstico médico asistido por inteligencia artificial.

  Para esta tarea, existen métricas ampliamente aceptadas que permiten medir distintos aspectos del comportamiento del modelo. Entre las más utilizadas se encuentran la precisión (precision), la recuperación (recall) y la puntuación F1 (F1

score). Cada una ofrece una perspectiva particular sobre los aciertos y errores del sistema.

La precisión mide la proporción de predicciones positivas que fueron efectivamente correctas, es decir, cuántas de las instancias que el modelo clasificó como positivas realmente pertenecían a esa clase. La recuperación, por su parte, indica cuántas de las instancias positivas reales fueron detectadas por el modelo. Finalmente, la puntuación F1 representa el promedio armónico entre precisión y recuperación, siendo especialmente útil en casos donde se requiere un equilibrio entre ambas métricas.

En el marco de los servicios cognitivos de Microsoft, estas métricas son empleadas como parte del proceso de evaluación de modelos personalizados de clasificación de texto. Según la documentación oficial de Azure AI Services, "la precisión mide qué proporción de las predicciones positivas del modelo fue correcta, mientras que la recuperación indica cuántas de las instancias positivas reales fueron identificadas por el modelo". Asimismo, se destaca que la puntuación F1 resulta especialmente útil en casos donde las clases están desbalanceadas o cuando las consecuencias de los errores son significativas (Métricas de evaluación, 2024).

Estas métricas no solo permiten analizar el rendimiento general de un modelo, sino que también orientan la toma de decisiones respecto al ajuste de parámetros, la selección de datos de entrenamiento y la elección de algoritmos. Su correcta aplicación asegura que los modelos desarrollados no solo sean técnicamente robustos, sino también confiables en contextos reales.

Un aspecto central en la evaluación de modelos de clasificación es el uso adecuado de métricas que reflejen fielmente su rendimiento. Tradicionalmente, la precisión, la recuperación y la puntuación F1 han sido las métricas más utilizadas para este fin, debido a su capacidad para capturar la relación entre los verdaderos positivos y los

errores cometidos por el modelo. Sin embargo, en el trabajo *"Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation"* (Ward Powers, 2011), señala que estas métricas, aunque útiles, no siempre ofrecen una visión completa, especialmente cuando se aplican en contextos donde los datos están desbalanceados o donde el costo de los errores varía significativamente. En su análisis, el autor propone el uso complementario de medidas como la informedness y la markedness, que tienen en cuenta tanto la sensibilidad como la especificidad del modelo, proporcionando una evaluación más equilibrada. Esta propuesta amplía el espectro de análisis y evidencia la necesidad de seleccionar métricas adecuadas según el contexto específico del problema.

De manera complementaria, el trabajo *"Estandarización de métricas de rendimiento para clasificadores Machine y Deep Learning"* (Borja-Robalino, Monleón-Getino, & Rodellar, 2020), aborda de forma sistemática la estandarización de métricas de evaluación en clasificadores aplicados a datos desbalanceados, lo cual resulta particularmente relevante para este trabajo, enfocado en las métricas recall, F1-score y precisión. A diferencia de enfoques tradicionales que priorizan el uso de accuracy, este estudio pone de manifiesto las limitaciones de dicha métrica en contextos multiclase con fuerte desequilibrio, proponiendo en su lugar un enfoque que incorpora medidas más sensibles como el índice Kappa. En consecuencia, a través de una herramienta diseñada en R, los autores realizan un análisis comparativo de varias métricas bajo distintos escenarios de reducción muestral. Cabe destacar que, dentro de este análisis, el F1-score se posiciona como una métrica intermedia robusta (más informativa que la precisión, aunque menos que el índice Kappa), lo que respalda su utilidad en la evaluación de modelos en contextos similares al presente estudio. Así, los resultados de esta investigación refuerzan la necesidad de adoptar métricas ajustadas al tipo de datos,

especialmente cuando se trata de conjuntos con clases minoritarias o distribuciones asimétricas.

Capítulo 3

Hipótesis

Y aquí estoy yo, tratando de hacer lo mejor, como si fuera mi Magnus opus, tratando de que estos momentos de lucidez se conviertan en un texto que quede como legado de lo que he imaginado.

Fuente: Autor

3 Hipótesis

La implementación de un modelo de software basado en Multiview Learning (ML) para la clasificación de imágenes mejorará la precisión en la identificación del tipo de imágenes consumidas, gracias a la integración de múltiples representaciones de cada imagen que capturan tanto características visuales como contextuales. Además, la retroalimentación obtenida mediante la interacción con sistemas de visualización avanzados permitirá optimizar continuamente el modelo, ajustando y refinando la clasificación en función de patrones dinámicos y preferencias del usuario.

Capítulo 4

Metodología

Poder crear algo de la propia imaginación, pensarlo, trabajarlo, soñarlo, idealizarlo, construirlo, volverlo a platear, fracasar y volver a crear es parte de la naturaleza de una mente creativa, y el hecho de definir un proyecto es tan solo el primer paso de una larga travesía con uno mismo.

Fuente: Autor

4 Metodología

Para la realización del proyecto de investigación se basó un enfoque metodológico cuantitativo, ya que se busca diseñar, implementar y evaluar un modelo de software orientado a la clasificación precisa de imágenes mediante el uso de Multiview Learning (MVL). El enfoque cuantitativo permite validar la hipótesis mediante mediciones sobre la precisión del modelo, esto enfocado en la pertinencia del contenido presentado y la experiencia del usuario. Además, se adoptan elementos del enfoque experimental al realizar pruebas controladas con prototipos funcionales del modelo en entornos inmersivos.

Al igual que en el marco referencial presentado anteriormente, se debe plantear una metodología sobre la cual se desarrolle la investigación, haciendo uso de las técnicas del método científico y partiendo desde aspectos deductivos que se puedan inferir de las temáticas con las respectivas características que se están abordando, así como inductivos que se basan los conocimientos que son tratados en el marco referencial que se planteó para el proyecto. Para ello estos métodos se ven descritos en las siguientes fases los cuales se verán de forma intrínseca desarrollados en el presente documento:

4.1 Definición y contextualización

En esta fase se definen los métodos que son utilizados para extracción de información relacionada con las imágenes además de definir los algoritmos de clasificación de imágenes a partir de técnicas que estén relacionadas con el MVL, esto con la finalidad de establecer la fundamentación teórica y práctica para la identificación de los aspectos más relevantes de las imágenes de contenido que se tendrán en cuenta para el desarrollo del proyecto, por tal motivo se realiza:

- *Los métodos empíricos:*

  Para el proyecto se plantea la formulación de experimentos de manejo de algoritmos de extracción, identificación, aprendizaje y clasificación de información relacionada con imágenes. Esto con la finalidad de realizar una observación de los resultados que se obtengan sobre un conjunto de imágenes de diverso tipo, de esta

forma se logra recolectar los metadatos de las imágenes, los cuales servirán para caracterizar las imágenes bajo vista de forma independiente y así establecer mediante medición cuales son los algoritmos que se utilizaron para obtener mejores resultados en el tratamiento de las imágenes.

- *Los métodos estadísticos:*

  Al momento de generar información a partir de las imágenes se busca establecer patrones que sean semejantes en contenido, para así implementar modelos estadísticos que ayuden a que el modelo de machine learning mejore su forma de analizar las imágenes. También, se utilizará para establecer las muestras entre las diversas bases de datos que se tengan disponibles.

4.2 Modelamiento del Sistema Multi View Learning (MVL)

- *Métodos teóricos:*
  - *Análisis y síntesis:* Se debe realizar la separación de cada uno de los algoritmos que se usaran para establecer las vistas. Para ello se hace un análisis basado en la bibliografía en el cual se tengan diversos métodos de tratamiento de imágenes para así establecer cuáles podrían usarse y que posibles resultados de esa transformación se puedan obtener. Ya con la información de cada uno de los algoritmos, se procederá a establecer las relaciones existentes entre los grupos de datos bajo una misma transformación y otra entre diferentes tipos de transformaciones, pero con el mismo origen de datos, para así lograr sintetizar y clasificar la información que posteriormente será visualizada.
  - *Hipotético deductivo:* Este método se implantará cuando a partir de un conjunto de casos particulares se pueda hallar patrones no establecidos previamente y que sirvan para describir una característica de las imágenes que se utilizaron para generar estos casos específicos. De este modo, al realizar esta deducción de los datos obtenidos se puede generalizar en una vista y que esta a su vez sirva para la clasificación de dichas imágenes, dando así paso de lo particular a lo general.

- *Modelación:* Para el manejo de la modelación del sistema de Multiview learning se deben establecer las etapas para el desarrollo de un proyecto de desarrollo de software.
    - *Análisis de requerimientos:*

    Establecer los requerimientos necesarios para el desarrollo del modelo de software que implemente la clasificación basada en MVL.
    - *Diseño y arquitectura:*

      Utilización de patrones arquitecturales y de desarrollo para el diseño de un modelo que integre la solución al problema planteado y que satisfaga los requerimientos planteados.

      Definición de tecnologías que se usan para este tipo de manejo del MVL y así establecer cuáles son las que más se adaptan a las necesidades de manejo de algoritmos para imágenes y con ello definir la arquitectura conceptual y matemática del modelo.
    - *Desarrollo:*

      Utilización de prácticas de código limpio para el desarrollo de la solución diseñada.
    - *Pruebas:*

      Establecimiento de escenarios de pruebas a partir de conjuntos conocidos de imágenes para la verificación del sistema de MVL según los parámetros establecidos en los requerimientos.
    - *Despliegue:*

      Generación de la forma de exponer la funcionalidad para los sistemas de visualización.

4.3 Modelamiento del sistema integrados de visualización

- *Modelación:* Para modelar la integración de los sistemas de visualización con el sistema de MVL, es necesario establecer las etapas para el desarrollo de un proyecto de software:
    - *Análisis de requerimientos:*

Establecer los requerimientos necesarios para la integración de cada uno de los dos sistemas existentes.

- *Diseño y arquitectura:*

Utilización de patrones arquitecturales y de desarrollo, para el diseño de un modelo que permita que cada una de las soluciones se desempeñen con la otra en conjunto y así generar un valor agregado a la experiencia del usuario.

- *Desarrollo:*

Utilización de prácticas de código limpio para el desarrollo de la solución diseñada.

- *Pruebas:*

Establecimiento de escenarios de pruebas a partir de la experiencia de los usuarios de los sistemas de visualización.

- *Despliegue:*

Generación de la integración total de ambos sistemas como si fueran uno.

4.4 Pruebas, Análisis de resultados y conclusiones

- *Los métodos estadísticos*:

  Se utilizará la estadística exploratoria y descriptiva para la caracterización de los resultados que se obtuvieron en las pruebas, tanto a nivel de mejoramiento del aprendizaje que se realiza en el sistema de MVL junto con sus resultados, como de la mejora en la experiencia del usuario a la hora de utilizar el sistema.

- Métodos teóricos:

  - *Inducción y deducción:*

  Para esta fase la inducción y deducción se utilizará a la hora de ver los resultados que arrojaron los métodos estadísticos para definir cuáles son los casos específicos y particulares donde se reflejen si estos conceptos generales son válidos o no para la implementación que se está realizando.

Capítulo 5

Diseño del Modelo de MVL

He analizado tantos problemas, pero no logro resolverme.
He diseñado tantas cosas, pero no puedo diseñar mi vida.
He desarrollado tantas soluciones, pero ninguna soluciona lo que necesito.
He probado tantos escenarios, pero no estoy preparado para el siguiente escenario.
He concluido muchos proyectos, pero este no será el último que haré.

Fuente: Autor

5 Diseño del modelo MVL

5.1 Boceto inicial del diseño del Modelo

El presente capítulo de este trabajo busca describir de forma general el proceso de diseño del Modelo de clasificación de imágenes a partir de Multiview Learning (MVL), siendo esta la forma en la cual se aborda la necesidad de mejorar la clasificación de imágenes teniendo en cuenta múltiples aspectos tanto visuales como contextuales, y de esta manera brindar a los sistemas de visualización una clasificación más pertinente y con la posibilidad de recibir una retroalimentación por parte de los usuarios.

El modelo que se propone tiene como finalidad integrar las múltiples perspectivas de cada imagen (visuales, de objetos, semánticas y de interacción de los usuarios), con el propósito de enriquecer el proceso de clasificación. Por tal motivo, se detallarán cada una de las partes del modelo, su definición y el aporte que brindan al modelo general.

Con todo lo anterior se estableció un proceso general de cómo se esperaba el funcionamiento y por ende el modelo, a continuación, se presenta el listado de los pasos

1. Se obtienen las imágenes del repositorio
2. Se aplican las técnicas de tratamiento de imágenes (vistas)
3. Se guardan los resultados
4. Se combinan dichos resultados para la clasificación
5. Se presentan en el aplicativo
6. Se obtienen las retroalimentaciones

Con los pasos anteriores se generó un diagrama de secuencia en donde se establecían el flujo básico de la información en el sistema

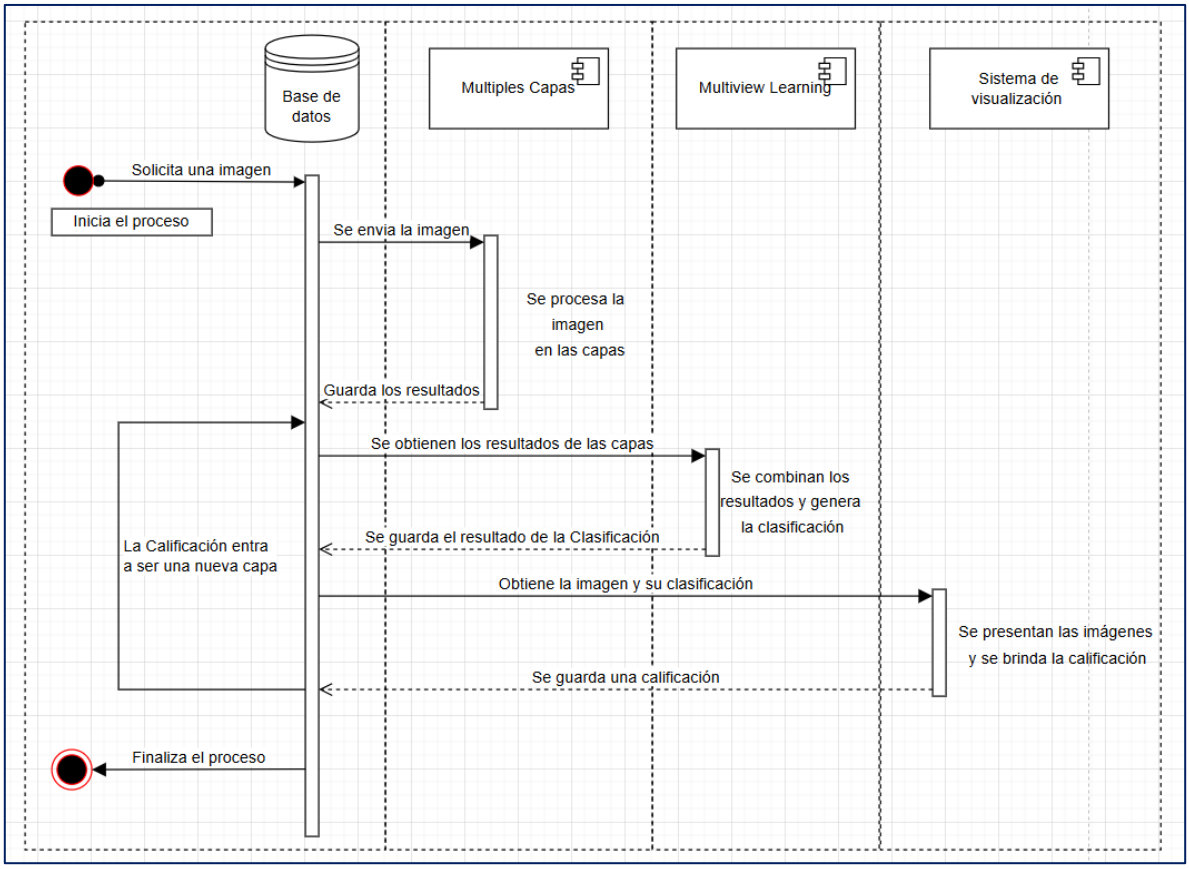

Ilustración 12: Diagrama de secuencia del modelo de Clasificación. Fuente: Autor

Con base en el diagrama anterior se tiene un boceto del funcionamiento y del cual se comienza a identificar cada uno de los módulos que requerirá el sistema, uno de ellos es el módulo de múltiples capas, en donde cada capa es una vista, y por ende se deben comenzar a definir cada una de las capas.

5.2 Definición de las capas del modelo

El *Multiview Learning* (MVL) es un enfoque del aprendizaje automático que explota la información complementaria proveniente de diferentes representaciones o "vistas" de los datos para mejorar la precisión y la generalización de los modelos predictivos. En este contexto de clasificación de imágenes lo que se busca es que el modelo pueda a partir de las diversas vistas tener una perspectiva para la clasificación y que al momento de combinar los resultados de cada cuna de estas vistas se pueda obtener un resultado de clasificación

más objetivo lo que se traduce en una clasificación más precisa y relevante para el sistema de visualización.

Para iniciar con la definición del trabajo se enmarcó un tema de estudio como lo es el conflicto armado en Colombia, lo cual es más complejo que la clasificación de imágenes enfocadas a un solo concepto como lo puede ser clasificación de animales o detección de personas que son temas usuales en la literatura.

A partir de este contexto se revisó la literatura que se ha depurado para el tema de MVL en este trabajo como base teórica, se comenzó explorando algoritmos que ya se han trabajado en algunos artículos, como por ejemplo se comenzó trabajando con algoritmos enfocados a contornos y formas (SIFT, HOG y LBP) tal como se hizo en el trabajo *"Multi-view learning for visual violence recognition with maximum entropy discrimination and deep features."* (Sun, Liu, & Mao, 2018), siendo este un referente, pero debido a que las imágenes son de múltiples la determinación de contornos no fue lo más acorde es porque se requerían de contornos muy específicos para una clasificación, es entonces donde si una imagen contenía una gran variación de bordes podía no clasificarla en sí. Cómo se puede ver en las siguientes imágenes se muestran patrones similares pero en menor escala de dos imágenes diferente esto se debe a la cantidad de bordes que se encontraron, pero esto no da una perspectiva clara de cual es una imagen que pueda estar relacionada con el conflicto (como lo puede estar la segunda imagen.

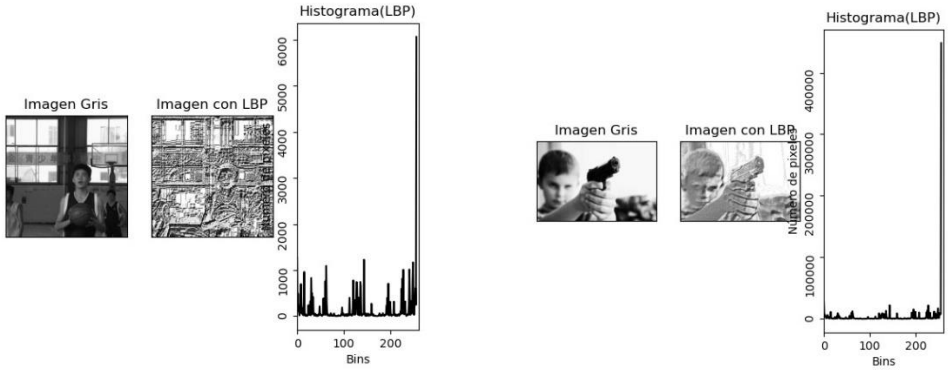

Ilustración 13: Resultados de la prueba de validación de la técnica LBP. Fuente: Autor

Es por ello que se optó por mirar técnicas más y que no estuvieran a tan bajo nivel, para poder identificar características más relevantes para el contexto por lo cual se continuó revisando la técnica más común y que ha tenido más variaciones en la literatura como lo son las Redes Neuronales Convolucionales.

5.2.1 Vista de Redes Neuronales Convolucionales

Esta vista se construye a partir de las representaciones generadas por una **red neuronal convolucional (CNN)**, que permite extraer patrones visuales de alto nivel de las imágenes. La CNN actúa como un codificador automático de la imagen, generando un vector de características que incluye información sobre textura, bordes, formas y estructura global del contenido visual.

El uso de las Redes Neuronales Convolucionales también se basa en las ventajas que se presentan, una de ellas es la posibilidad de utilizar en librerías que ya están desarrolladas y son especiales para el manejo de imágenes y flujo de redes neuronales, otra ventaja es la capacidad de realizar un proceso de entrenamiento de imágenes que permite tener un control del aprendizaje (aprendizaje supervisado) para lo cual es importante para la base visual general de esta capa y posteriormente realizar una validación con un set de datos más amplio.

5.2.2 Vista de Análisis de metadatos

Esta vista es una capa de tipo semántico en la cual se busca extraer las palabras clave relevantes del título y/o descripción y/o metadatos que pueda tener una imagen que esté almacenada en el sistema, esto a partir de la definición de un diccionario temático establecido para el contexto del caso de estudio.

El uso de esta información textual para ser transformada en vectores de frecuencia aporta contexto narrativo sobre el contenido representado en la imagen. Siendo esta una parte fundamental del modelo ya que una imagen puede tener una representación bastante sencilla pero el hecho de tener un texto que brinde una narrativa contextual ayuda a generar un valor agregado al momento de clasificar la imagen.

5.2.3 Vista de Detección de Objetos

Esta vista es de análisis visual del contenido de la imagen para la cual se estableció la técnica de YOLO **(You Only Look Once)** para detectar objetos específicos dentro de la imagen. Esto sirve para poder definir una serie de elementos concretos como como personas, armas, vehículos, paisajes, entre otros elementos, los cuales son particularmente relevantes en imágenes que puedan representar un contenido (que para el caso de estudio) es relevante.

Este tipo de técnica, al igual que las redes neuronales presenta muchas ventajas, entre las cuales se destaca la posibilidad de utilizar librerías que ya contienen esta técnica implementada, y la versatilidad de utilizar etiquetas para los objetos que se quieren ubicar permitiendo no solo detectar uno sino varios objetos dentro de una sola imagen, lo cual ayuda al sistema a tener un criterio objetivo sobre qué elementos componen la imagen a nivel de contenido.

5.2.4 Vista de Retroalimentación de los usuarios

Esta vista es para un análisis cuantitativo de la información que proviene de la interfaz que obtiene datos de un sistema de visualización, en donde se recibe la imagen y su valoración de cuan pertinente es esta para el usuario; estos datos se integran al modelo cuando se hace una reclasificación de una imagen ya que previamente ha sido procesada sin tener en cuenta esta retroalimentación. Cuando se integran, comienzan a hacer parte del proceso en donde entran a ser un vector de análisis para el sistema de Multiview Learning y así mejora la determinación del sistema de Multiview ya sea para compensarlo o para reafirmar la valoración generada.

5.3 Modelo de Multiview Learning

A partir de las consideraciones presentadas en este capítulo se establece el modelo base del sistema de clasificación basado en *Multiview Learning* (MVL)

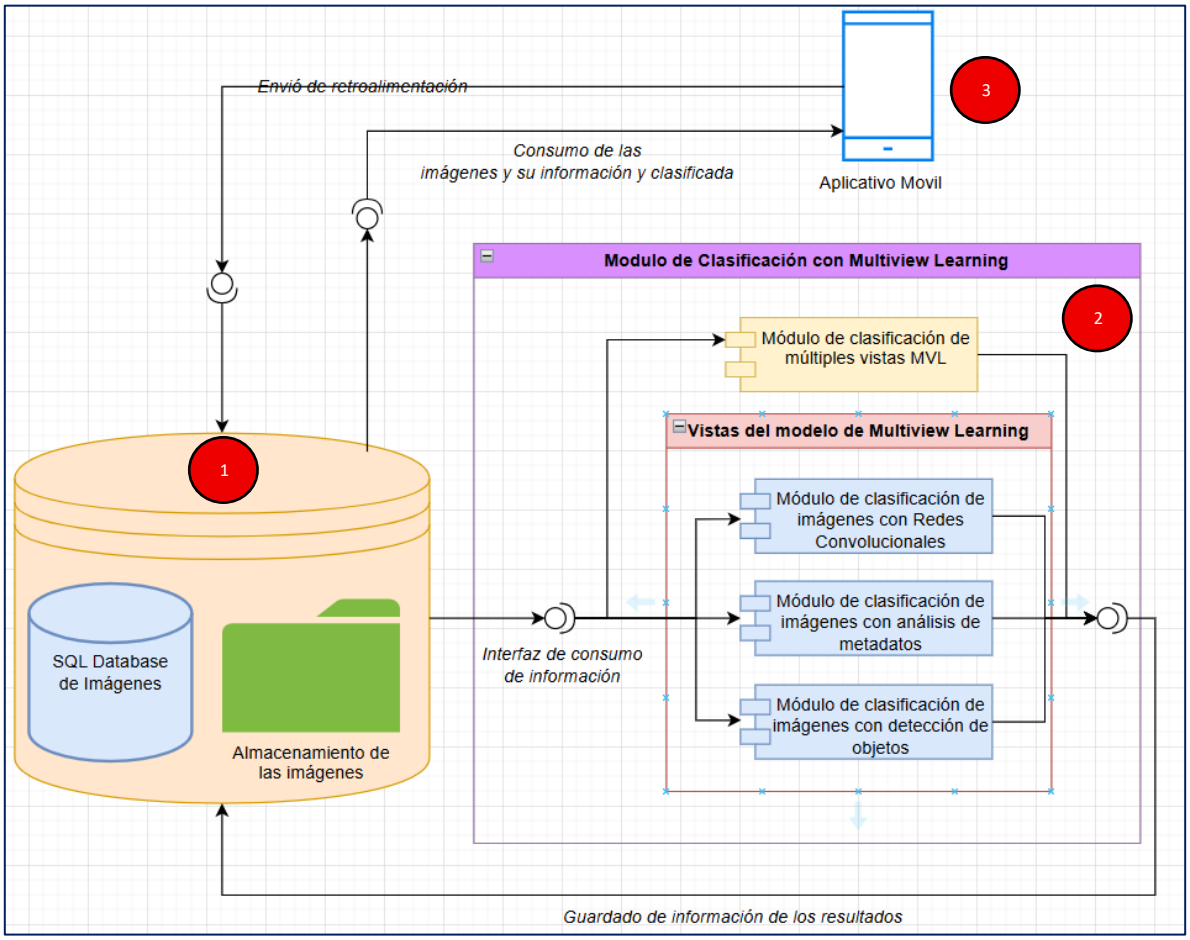

Ilustración 14: Diagrama de componentes del sistema de Clasificación de Imágenes en conjunto con un sistema de Visualización. Fuente: Autor

En este diagrama se representan los diferentes componentes que hacen que conforma el sistema de clasificación. A continuación, se detallan cada uno de ellos para su contextualización:

1. Base de datos: esta componente hace referencia a toda la capa de persistencia sobre la cual se almacenan los datos y las imágenes que se utilizaran para el caso de estudio. Se encuentra conformada por dos elementos: una base de datos SQL para

la información de las imágenes y demás elementos persistentes del sistema, y un almacenamiento de imágenes que contiene los objetos a trabajar.

2. Módulo de Clasificación con Multiview Learning: Este módulo abarca todo el sistema de clasificación de Multiview en el cual se encuentran los siguientes componentes
    a. Vistas del modelo de Multiview Learning: Son las capas que se han establecido para que se haga una preclasificación en cada una de ellas. Como se planteó en los ítems de las vistas se manejan los tres componentes para una clasificación inicial
        i. Módulo de Clasificación de Imágenes con Redes Convolucionales
        ii. Módulo de Clasificación de Imágenes con análisis de metadatos
        iii. Módulo de Clasificación de Imágenes con detección de Objetos
    b. Módulo de Clasificación de múltiples Vistas MVL: Este es el componente que reúne la información almacenada de cada una de las vistas para poder condensarlas y tomar la determinación final de la clasificación.
3. Sistema de visualización (aplicación móvil): es el sistema trabajado en el documento *"DISEÑO DE UN METAMODELO DE SOFTWARE PARA EL DESARROLLO DE APLICACIONES INMERSIVAS"* (Sánchez Cruz, 2025) con el cual se colabora para la presentación del contenido ya clasificado y posterior recopilación de las valoraciones por parte de los usuarios, con el fin de completar la retroalimentación y de esta manera tener información como parte de la vista de "Retroalimentación de los usuarios" la cual está planteada en el ítem 5.2.4.

Capítulo 6

Implementación del Modelo

Cientos de días y miles de noches,
una certeza e incontables incertidumbres,
una calma con constante estrés,
la satisfacción enmarcada en la depresión,
En fin todo lo trabajado, todo lo que inspiró
para calmar esta batalla conmigo mismo.

Fuente: Autor

6 Implementación del modelo

6.1 Introducción

En este capítulo del proyecto se describe cada uno de los elementos establecidos desde el capítulo anterior “Diseño del modelo MVL”. Se detalla no solo la construcción de cada una de las vistas definidas sino también aspectos relevantes para su implementación, esto con la finalidad de mostrar desde un punto de vista de la arquitectura del sistema cuales son las funcionalidades, entornos, componentes y tecnologías que se utilizaron para el desarrollo e implementación de este modelo, también se presentan aspectos claves como la base de datos y su estructuración.

Y por último lograr evidenciar en este capítulo el proceso técnico que sustenta esta solución propuesta, mostrando su viabilidad práctica, su capacidad de extensibilidad y su alineación con los objetivos planteados en este trabajo de investigación.

6.2 Estructuración de la Base de datos

Para este proceso de estructuración de base de datos se partió estableció para cumplir con tres objetivos principales que el sistema requiere tanto para el sistema de clasificación de imágenes como para el sistema de visualización que hace uso del el, los tres objetivos base son:

1. Almacenar la información de la meta data de las imágenes en una base de datos SQL
2. Almacenar información de los usuarios, recorridos y resultados de la experiencia
3. Almacenar los resultados de las capas y del módulo de Multiview Learning

Con dichos objetivos se generó un diagrama de base de datos que lograra cubrir estas necesidades

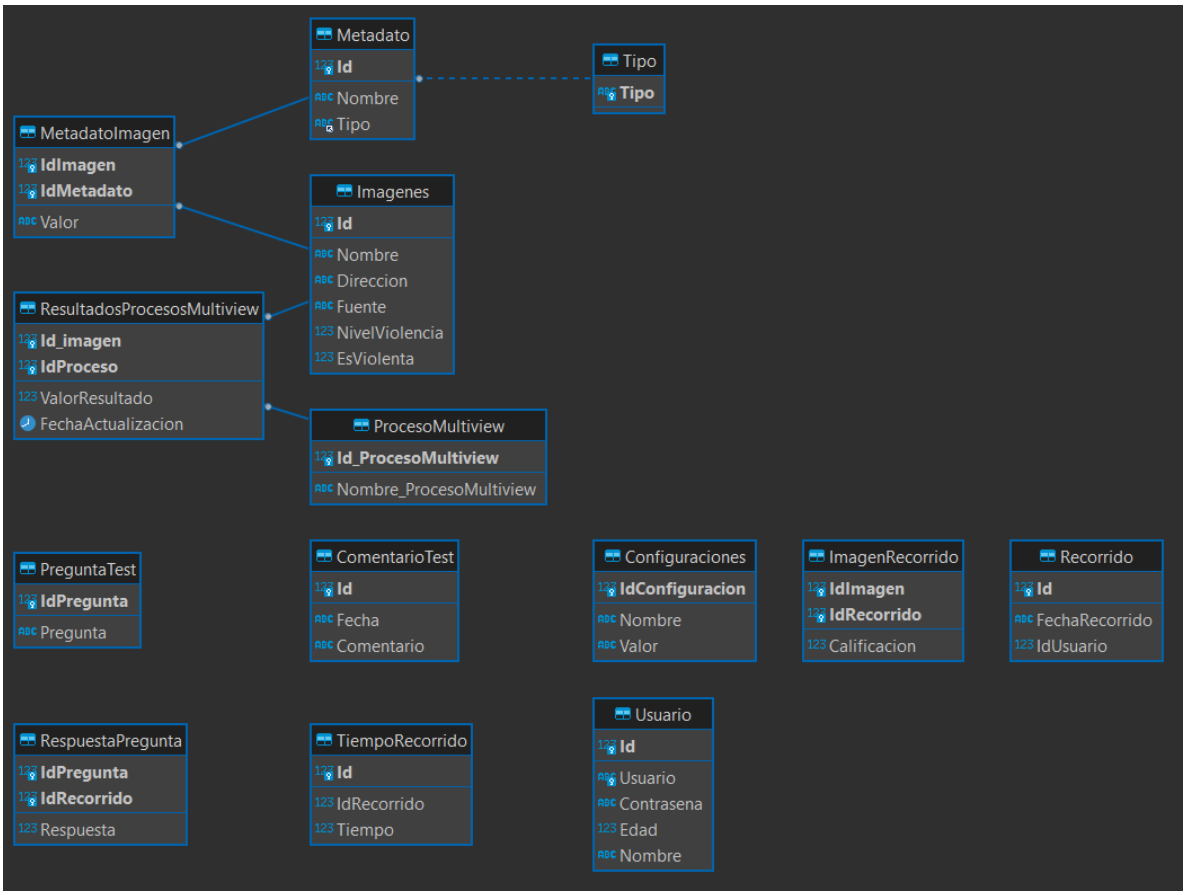

Ilustración 15: Diagrama Entidad Relación de la base de datos utilizadas para el proyecto. Fuente: Autor

En este diagrama se destacan tablas que solo pertenecen al sistema de visualización como lo son las tablas de Usuario, Recorrido, Configuraciones, etc. Y sobre todo para el caso de este proyecto caben resaltar las tablas de Imagen, Metadato, Proceso Multiview y Resultados Procesos Multiview, para las cuales a continuación se hace una explicación de la estructuración de esta base de datos.

6.2.1 Almacenamiento de la información de las imágenes

Partiendo de estos objetivos también se tuvo en cuenta la premisa de la delimitación del caso de estudio enfocado al conflicto armado, para lo cual se realizó un trabajo de investigación y recopilación de imágenes que estuvieran relacionadas específicamente con este caso de estudio.

Como se explicó en el diseño del modelo MVL la base de datos tiene dos componentes que son la base de datos SQL y el repositorio donde se almacenan las imágenes, para lo cual es

necesario que cada imagen tenga la información tanto en la base de datos como en si la imagen que se recopiló, lo cual genero un amplio set de datos relacionados con el conflicto armado.

Esta información fue recopilada de diversos orígenes de información para los cuales se almacena su respectivo origen y se le atribuye la autoría a su respectivo autor debido a que este proyecto es netamente académico y las imágenes fueron usadas solo para este enfoque de investigación.

repositorionuevo | ResultadosProcesosMultiview | <repositorionuevo> Script | ProcesoMultiview | Imagenes | ImagenRecorrido

Propiedades | Datos | Diagrama ER | repositorionuevo | Databases | repositorionuevo

imagenes | Enter a SQL expression to filter results (use Ctrl+Space)

	Id	Nombre	Direccion	Fuente
36	149	0. Conflicto armado interno de Colombia - Wikipedia, .png	Violentas/conflicto armado en colombia/149.pr	https://es.wikipedia.org/wiki/Conflict
37	150	1. Por qué el conflicto armado en Colombia necesita d.jpeg	Violentas/conflicto armado en colombia/150.jp	https://www.elpais.com.co/judicial/p
38	151	2. Conflicto armado en Colombia_ factores, actores y .jpeg	Violentas/conflicto armado en colombia/151.jp	https://www.elmundo.com/portal/nc
39	152	3. Escenarios naturales del conflicto armado en Colom.jpeg	Violentas/conflicto armado en colombia/152.jp	https://agronegocios.uniandes.edu.c
40	153	4. Casi 600000 víctimas del conflicto colombiano fuer.jpeg	Violentas/conflicto armado en colombia/153.jp	https://www.semana.com/pais/articu
41	154	5. El conflicto armado interno en Colombia dejó más d.jpeg	Violentas/conflicto armado en colombia/154.jp	https://www.aa.com.tr/es/mundo/el-
42	155	6. El alto costo económico y social del conflicto arm.jpeg	Violentas/conflicto armado en colombia/155.jp	https://www.elheraldo.co/economia/
43	156	7. Colombia enfrenta consecuencias por no reconocer s.jpeg	Violentas/conflicto armado en colombia/156.jp	https://latinamericanpost.com/es/31
44	157	8. Colombia ha perdido 7 millones de años de vida por.jpeg	Violentas/conflicto armado en colombia/157.jp	https://www.kienyke.com/colombia/
45	158	9. Los cinco conflictos armados que persisten en Colo.jpeg	Violentas/conflicto armado en colombia/158.jp	https://www.rcnradio.com/colombia,
46	159	11. Los 7 puntos claves para entender el conflicto arm.jpeg	Violentas/conflicto armado en colombia/159.jp	https://www.infobae.com/america/a
47	160	12. El impacto psicosocial de los niños víctimas del c.jpeg	Violentas/conflicto armado en colombia/160.jp	https://www.semana.com/educacion
48	161	13. Más de 16000 menores fueron reclutados durante el .jpeg	Violentas/conflicto armado en colombia/161.jp	https://www.elheraldo.co/colombia/
49	162	14. Cifra de menores víctimas del conflicto armado en .jpeg	Violentas/conflicto armado en colombia/162.jp	https://www.elpais.com.co/proceso-
50	163	15. Informe_ Más de 260 mil muertos en conflicto armad.jpeg	Violentas/conflicto armado en colombia/163.jp	https://www.hispantv.com/noticias/c
51	164	17. La difícil realidad del conflicto armado en Colomb.jpeg	Violentas/conflicto armado en colombia/164.jp	https://www.icrc.org/es/document/c
52	165	18. Conflicto armado en colombia.jpeg	Violentas/conflicto armado en colombia/165.jp	https://es.slideshare.net/ManuelVela
53	166	19. Las voces de los militares en el conflicto armado .jpeg	Violentas/conflicto armado en colombia/166.jp	https://www.unisabana.edu.co/portal
54	167	20. Para el CICR, en Colombia hay 5 conflictos armados.jpeg	Violentas/conflicto armado en colombia/167.jp	https://www.elcolombiano.com/colo
55	168	21. Conflicto Armado Colombia_ imágenes, fotos de stoc.jpeg	Violentas/conflicto armado en colombia/168.jp	https://www.shutterstock.com/es/sea
56	169	22. La Fiscalía reveló cifras del conflicto armado en .jpeg	Violentas/conflicto armado en colombia/169.jp	https://m.elfrente.com.co/index.php?
57	170	23. Conflicto armado en Colombia_ tema, información y .jpeg	Violentas/conflicto armado en colombia/170.jp	https://cnnespanol.cnn.com/tag/cont
58	171	24. Colombia – Registro de fallecimiento de menores po.jpeg	Violentas/conflicto armado en colombia/171.jp	https://www.diariojuridico.com/colo
59	172	25. Los 7 puntos claves para entender el conflicto arm.jpeg	Violentas/conflicto armado en colombia/172.jp	https://www.infobae.com/america/a
60	173	26. Claves para leer el nuevo rumbo del conflicto arma.jpeg	Violentas/conflicto armado en colombia/173.jp	https://latinta.com.ar/2019/08/claves
61	174	27. Quieren reactivar el conflicto armado colombiano_ .png	Violentas/conflicto armado en colombia/174.pr	https://www.las2orillas.co/quieren-re
62	175	28. Colombia_ el conflicto armado dejó unos 262000 mue.jpeg	Violentas/conflicto armado en colombia/175.jp	https://apnews.com/01854d898cfd49
63	176	29. Violencia por conflicto armado en Colombia _causó .jpeg	Violentas/conflicto armado en colombia/176.jp	https://mundo.sputniknews.com/202
64	177	30. Conflicto armado colombiano en su nivel más bajo c.jpeg	Violentas/conflicto armado en colombia/177.jp	https://www.radionacional.co/noticia

Ilustración 16: Muestra de la tabla de imágenes con el título, ubicación y fuerte de las imágenes. Fuente: Autor

Posteriormente se realizó la recopilación de información de imágenes que son completamente ajenas a la temática del conflicto armado en Colombia, siendo este un contrapeso para la validación de las clasificaciones. Al ser un proceso de extracción de imágenes que pueden ser de cualquier tipo de imágenes se utilizó un banco de imágenes llamado Flickr el cual proporciona una Api la cual fue utilizada mediante un scrapper para la obtención de las imágenes y su respectiva metadata que cuenta con uso libre.

```
Users > KevinCamiloDiazMonje > OneDrive - SKITCONSULTING LTDA > Personal > Universidad > Maestria > Tesis 2 > Proyecto de tesis > Descargar imagenes > flickr_scraper > flickr_s
1   # Generated by Glenn Jocher (glenn.jocher@ultralytics.com) for https://github.com/ultralytics
2
3   import argparse
4   import os
5   import time
6   import pandas as pd
7
8   from flickrapi import FlickrAPI
9
10  from utils.general import download_uri
11
12  key = '2[redacted]'  # Flickr API key https://www.flickr.com/services/apps/create/apply
13  secret = '[redacted]
14
15
16  def get_urls(search='gatos', n=20, download=False):
17      t = time.time()
18      flickr = FlickrAPI(key, secret)
19      license = ('9')  # https://www.flickr.com/services/api/explore/?method=flickr.photos.licenses.getInfo ------- public domain
20      photos = flickr.walk(text=search,  # http://www.flickr.com/services/api/flickr.photos.search.html
21                           extras='url_o',
22                           per_page=500,  # 1-500
23                           license=license,
24                           sort='relevance')
```

Ilustración 17: Vista inicial del código del scrapper. Fuente: Autor

Esto generó una cantidad de imágenes que no estuvieran relacionadas con el contexto del caso de estudio como se presenta a continuación

repositorionuevo | ResultadosProcesosMultiview | <repositorionuevo> Script | ProcesoMultiview | Imagenes | ImagenRecorrido

Propiedades | Datos | Diagrama ER | repositorionuevo | Databases | repositorionuevo | Tables

Imagenes | Enter a SQL expression to filter results (use Ctrl+Space)

	Id	Nombre	Direccion	Fuente
3258	3.378	19. Volcano Etna at sunset - Creative Commons by gnuckx	NoViolentas/agricultura/3378.jpg	https://www.flickr.com/photos/3440
3259	3.379	20. 038 gnuckx Etna_Volcano-Italy-castielli_CC0-HQ	NoViolentas/agricultura/3379.jpg	https://www.flickr.com/photos/3440
3260	3.380	1. Amistad	NoViolentas/amistad/3380.jpg	https://www.flickr.com/photos/6251
3261	3.381	2. Parque de la Amistad	NoViolentas/amistad/3381.jpg	https://www.flickr.com/photos/8105
3262	3.382	3. La Amistad Bed & Breakfast	NoViolentas/amistad/3382.jpg	https://www.flickr.com/photos/6251
3263	3.383	4. 472 - LESSER NIGHTHAWK (10-21-2015) lake amistad, val verde	NoViolentas/amistad/3383.jpg	https://www.flickr.com/photos/8101
3264	3.384	5. AMISTAD	NoViolentas/amistad/3384.jpg	https://www.flickr.com/photos/1531
3265	3.385	6. amistad	NoViolentas/amistad/3385.jpg	https://www.flickr.com/photos/1356
3266	3.386	7. Amistad.	NoViolentas/amistad/3386.jpg	https://www.flickr.com/photos/1855
3267	3.387	8. Amistad School Visit	NoViolentas/amistad/3387.jpg	https://www.flickr.com/photos/4735
3268	3.388	9. Amistad School Visit	NoViolentas/amistad/3388.jpg	https://www.flickr.com/photos/4735
3269	3.389	10. Amistad School Visit	NoViolentas/amistad/3389.jpg	https://www.flickr.com/photos/4735
3270	3.390	11. Amistad School Visit	NoViolentas/amistad/3390.jpg	https://www.flickr.com/photos/4735
3271	3.391	12. Amistad School Visit	NoViolentas/amistad/3391.jpg	https://www.flickr.com/photos/4735
3272	3.392	13. Parque de la Amistad	NoViolentas/amistad/3392.jpg	https://www.flickr.com/photos/8105
3273	3.393	14. Parque de la Amistad	NoViolentas/amistad/3393.jpg	https://www.flickr.com/photos/8105
3274	3.394	15. Parque de la Amistad	NoViolentas/amistad/3394.jpg	https://www.flickr.com/photos/8105
3275	3.395	16. Parque de la Amistad	NoViolentas/amistad/3395.jpg	https://www.flickr.com/photos/8105
3276	3.396	17. Parque de la Amistad	NoViolentas/amistad/3396.jpg	https://www.flickr.com/photos/8105
3277	3.397	18. Parque de la Amistad	NoViolentas/amistad/3397.jpg	https://www.flickr.com/photos/8105
3278	3.398	19. Parque de la Amistad	NoViolentas/amistad/3398.jpg	https://www.flickr.com/photos/8105
3279	3.399	20. Parque de la Amistad	NoViolentas/amistad/3399.jpg	https://www.flickr.com/photos/8105
3280	3.400	1. Amor	NoViolentas/amor/3400.jpg	https://www.flickr.com/photos/6251
3281	3.401	2. Amor	NoViolentas/amor/3401.jpg	https://www.flickr.com/photos/6251
3282	3.402	3. Amor	NoViolentas/amor/3402.jpg	https://www.flickr.com/photos/6251
3283	3.403	4. Amor	NoViolentas/amor/3403.jpg	https://www.flickr.com/photos/6251
3284	3.404	5. Amor paterno filial	NoViolentas/amor/3404.jpg	https://www.flickr.com/photos/6251
3285	3.405	6. Amor	NoViolentas/amor/3405.jpg	https://www.flickr.com/photos/6251
3286	3.406	7. Amer, Jaipur	NoViolentas/amor/3406.jpg	https://www.flickr.com/photos/4351

Ilustración 18: Muestra de la tabla de imágenes con solo imágenes obtenidas del scrapper. Fuente: Autor

Una vez definido cómo se obtuvieron las imágenes se agregaron los metadatos y su respectiva relación a cada una de las imágenes para así obtener una información más segmentada de las imágenes, esto permitiendo agregar metadatos al sistema y en casos en donde no se tengan todos los metadatos de una imagen solo se agreguen los que se tengan disponibles. A continuación, se muestra un ejemplo de la tabla de Metadato

Metadato × repositorionuevo Resu

Propiedades Datos Diagrama ER

Metadato Enter a SQL expression to filter re

Grilla Texto

	Id	Nombre	Tipo
1	1	Lugar	Texto
2	2	Fecha captura	Fecha
3	4	Autor	Texto
4	5	B/N	Texto
5	6	Violenta	Texto
6	7	PalabrasClaves	Texto
7	8	Objetos	Texto
8	9	Descripcion	Texto
9	10	Sala	Texto
10	11	URL	Texto
11	12	Contexto	Texto

Ilustración 19: Tabla de Metadato. Fuente: Autor

A continuación, se presenta la tabla MetadatoImage (la cual es la tabla que integra la información de la imagen y sus metadatos)

MetadatoImagen × Metadato repositorionuevo ResultadosProceso... <repositorionuevo...

Propiedades Datos Diagrama ER repositorionuev

MetadatoImagen Enter a SQL expression to filter results (use Ctrl+Space)

Grilla Texto

	IdImagen	IdMetadato	Valor
1	29	1	Trujillo (Valle del Cauca)
2	29	2	01/01/2008
3	29	4	Jesús Abad Colorado
4	29	5	No
5	29	6	No
6	29	7	Marcha, peregrinación, monumento
7	29	8	Personas, Fotos
8	29	9	Marcha de familiares de las victimas
9	29	10	1
10	29	11	http://centrodememoriahistorica.gov.co/micrositios/expo_itinerante/
11	29	12	Trujillo
12	30	1	Trujillo (Valle del Cauca)
13	30	4	Jesús Abad Colorado
14	30	5	No
15	30	6	No
16	30	7	Grabados
17	30	8	Jeroglíficos
18	30	9	Multiples grabados de personas
19	30	10	1
20	30	11	http://centrodememoriahistorica.gov.co/micrositios/expo_itinerante/
21	30	12	Trujillo

Ilustración 20: Tabla MetadatoImagen. Fuente: Autor

6.2.1.1 Conexión a la Base de datos

Para poder hacer uso desde las vistas que se trabajan se realizó un script que se usa para la conexión a la base de datos tanto para poder hacer la consulta de las imágenes como para el guardado de los resultados generados por cada una de las vistas

```
cnn_violencia.py 8 ●   ConexionDB.py 5 ×   prepare_dataset.py ●   webscraping.py 6 ●
C: > Users > KevinCamiloDiazMonje > OneDrive - SKITCONSULTING LTDA > Personal > Universidad > Maestria > Tesis 2 > Proyecto de tesis > 2. Analisis
1   from datetime import datetime
2   import sqlite3
3   import requests
4   import sqlalchemy
5   from sqlalchemy.ext.automap import automap_base
6   from sqlalchemy.orm import Session
7   from sqlalchemy import create_engine
8
9   #Conexión a una BD local
10  class conexionBD:
11      def __init__(self, ruta):
12          self.ruta = ruta
13          self.conex=sqlite3.connect(ruta)
14          self.cursorObj=self.conex.cursor()
15
16      def close(self):
17          self.conex.close()
18
19   #   def obtenerMetadatosImagenes(self, idmetadato):
20   #       return self.cursorObj.execute("SELECT * from MetadatoImagen where IdMetadato ='"+idmetadato+"'")
21
22  #Conexión a la api imagenes : http://35.233.219.175/api/imagenes
23  class conexionApi:
24      def __init__(self,ruta):
25          self.ruta =ruta
```

```
49  #Conexión BD Mysql
50  class conexionBDMSQL:
51      def __init__(self, ruta):
52          self.ruta = ruta
53
54          self.Base = automap_base()
55
56          # engine, suppose it has two tables 'user' and 'address' set up
57          engine = create_engine(ruta)
58
59          # reflect the tables
60          self.Base.prepare(engine, reflect=True)
61
62          self.session = Session(engine)
63
64      def guardarResultadoProceso(self,idImagen, valorResultado):
65          ResultadosProcesosMultiview = self.Base.classes.ResultadosProcesosMultiview
66          var = self.session.merge(ResultadosProcesosMultiview(Id_imagen=idImagen,ValorResultado=valorResultado,IdProceso=2,FechaActuali
67          self.session.commit()
68
```

Ilustración 21: Visualización del código para la conexión y envió del resultado de cada vista. Fuente: Autor

6.2.1.2 Repositorio de Imágenes

Una vez establecida la estructura de los datos que se almacenan en la base de datos y que son propios de las imágenes que se van a trabajar se estableció el repositorio de imágenes en donde se almacenan las imágenes, inicialmente se establecieron en una segmentación de dos carpetas en donde se recopilaron las imágenes "Violentas" que hacen referencia a las imágenes del conflicto armado, y una carpeta de "NoViolentas" que hace referencia a las imágenes que fueron descargadas mediante el scrapper.

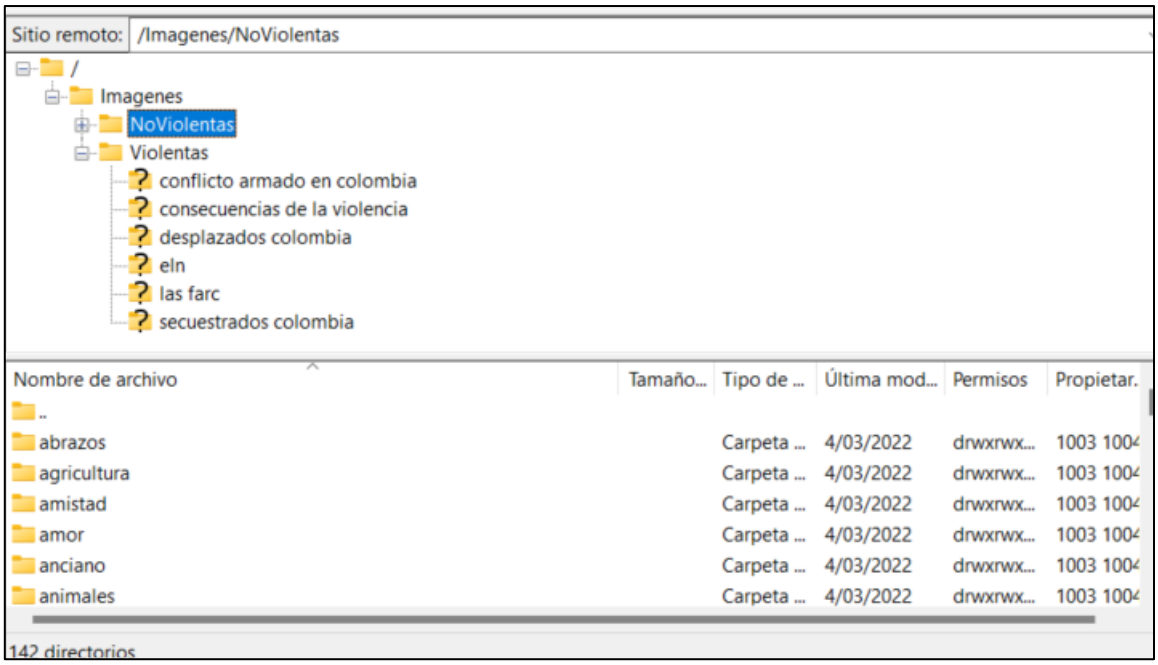

Ilustración 22: Visualización de la estructura del repositorio de imágenes. Fuente: Autor

Con todo lo anterior se logró establecer un conjunto de imágenes de 5719 imágenes distribuidas en entre las carpetas "Violentas" y "NoViolentas" y sus respectivas subcarpetas

A continuación, se presentan una pequeña muestra de las imágenes relacionadas con el contexto de conflicto armado en Colombia para dar un contexto de la diversidad de contenido que se presentan bajo este caso de estudio.

Ilustración 23: Imágenes del repositorio enfocadas al conflicto armado. Fuente: Autor

Como se puede apreciar en la imagen anterior las imágenes relacionadas con conflicto armado son variables en su forma de ser, entre las cuales se ven marchas, lugares desolados, paisajes abandonados, grupos armados, fotos de desaparecidos etc., lo cual hace que no sea un conjunto homogéneo para su clasificación como lo podría ser clasificar entre perros y gatos o entre personas y objetos.

Cabe anotar que si bien algunos símbolos nacionales o culturales, como la bandera de Colombia o las indumentarias típicas de ciertas regiones como “ponchos” o “sombreros”, han sido utilizados en contextos asociados a la violencia, su significado no es intrínsecamente violento. La semiótica indica que los símbolos son polisémicos y que su interpretación depende del contexto en el que aparecen. (Eco, 1976)

En relación con el sesgo derivado de la estructuración y segmentación de las imágenes que se manejan en el proyecto, tal como se describe en el artículo “A survey on bias in visual datasets” (Fabbrizzi, Papadopoulos, Ntoutsi, & Kompatsiaris, 2022), no existe un conjunto de datos libre de sesgo, dado que los datasets puede presentar cualquiera de los tipos de sesgo como:

- Sesgo de selección: al haber elegido el tema particular existe una sobre representación en los conjuntos de datos, pero a su vez es necesario para enmarcar

el caso puntual del conflicto armado en Colombia y que no se relacione con otro tipo de conflicto

- Sesgo de encuadre (Enfocarse en un tema en específico): al ser imágenes que no tienen una composición fotográfica estructurada los ángulos de las tomas, el contexto que rodea a la imagen principal, las personas, etc. pueden generar asociaciones que estén muy relacionados a ciertos estereotipos que hacen que las imágenes al momento de estructurar el dataset queden sesgadas, de igual manera es parte del trabajo ya que con el planteamiento lo que se busca es poder dar un valor desde diversos puntos de vista.
- Sesgo de etiquetado: Como suele ocurrir en los procesos de selección inicial de los datasets, se puede incurrir de forma inconsciente a un etiquetado errado en menor proporción (ya se automático o manual) ya que pueden contener etiquetas difusas, mal asignadas, o esto mismo se puede asignar sin una previa estructuración de los criterios de clasificación

En resumen, el sesgo que se presenta en el dataset, no se debe percibir como un incidente o limitante sino como un recurso intencional para poder abordar y acotar puntualmente el caso de estudio y permitiendo que se garantice un entrenamiento bastante especifico y coherente al contexto del conflicto armado.

Por último, se hace aclaración de que, en este trabajo, las imágenes se emplean con un propósito estrictamente académico y técnico, orientado a la representación visual y a la experimentación metodológica, sin pretender reforzar asociaciones negativas ni violentas. Por el contrario, al utilizar estos elementos se busca reconocer la riqueza cultural, la memoria histórica y simbólica del país, entendiendo que cualquier interpretación de violencia corresponde más a los contextos históricos de uso que al símbolo en sí mismo.

6.2.2 Almacenamiento de información de usuarios

Este aspecto va más relacionado a los usuarios y toda la información que requiere el sistema de visualización para su uso, como se mencionaba al inicio del capítulo esta base de datos realiza la tarea de persistencia de la información de los usuarios, de los recorridos que hace un usuario en el sistema de visualización y a su vez en dicho recorrido se almacena cuáles son los id's de las imágenes que se presentaron, y por último almacenando el resultado de las calificaciones.

6.2.3 Almacenamiento de información de resultados

Este objetivo de almacenar la información de los resultados va netamente ligado a la información que se almacena en cada una de las vistas de tal forma que cada una de las vistas puede funcionar de manera independiente y sobre la cantidad de imágenes que se requiera y de esta forma almacenar la valoración que cada una de las vistas brindo al momento de ejecutar su proceso propio de clasificación.

Para el manejo de esta información se trabajaron las siguientes tablas, primero la tabla de Procesos Multiview la cual muestra un id por cada capa para poder relacionarla con el resultado.

ProcesoMultiview | repositorionuevo | ResultadosProce...

Propiedades | Datos | Diagrama ER

ProcesoMultiview | Enter a SQL expression to filter results (use Ctrl+Sp

Grilla | Texto

	Id_ProcesoMultiview	Nombre_ProcesoMultiview
1	1	ProcesoRNC
2	2	ProcesoMetadato
3	3	ProcesoObjetos
4	4	ProcesoMultiview
5	5	ResultadoPersona

Ilustración 24: Tabla de ProcesoMultiview. Fuente: Autor

El resultado de cada vista por cada imagen se muestra en la tabla ResultadosProcesosMultiview

ResultadosProcesosMultiview × ProcesoMultiview repositorionuevo <repositor

Propiedades Datos Diagrama ER repositori

ResultadosProcesosMultiview Enter a SQL expression to filter results (use Ctrl+Space)

Grilla Texto

	Id_imagen	ValorResultado	IdProceso	FechaActualizacion
1	29	0,695	1	2023-06-07 05:04:27
2	29	0,2	2	2023-06-07 05:04:27
3	29	0	3	2023-10-11 19:13:14
4	29	1	4	2023-06-27 05:29:50
5	30	0,651	1	2023-06-07 05:04:27
6	30	0,75	2	2023-06-07 05:04:35
7	30	0	3	2023-10-11 19:13:14
8	30	1	4	2023-06-27 05:29:55
9	31	0,172	1	2023-06-07 05:04:27
10	31	0,2	2	2023-06-07 05:04:43
11	31	0	3	2023-10-11 19:13:33
12	31	0	4	2023-06-27 05:29:59
13	32	0,906	1	2023-06-07 05:04:27
14	32	0,333	2	2022-10-12 00:15:20
15	32	0	3	2023-10-11 19:13:14
16	32	1	4	2023-06-27 05:30:04

Ilustración 25: Tabla de ResultadosProcesosMultiview. Fuente: Autor

A modo de ejemplo se muestra que la imagen con el id=29, al hacer los procesos 1,2,3 y 4 que corresponden a los 3 procesos de las vistas definidas y al proceso de Multiview learning tiene un registro con la respectiva valoración que cada una de las capas brindó.

6.3 Arquitectura

A nivel de arquitectura se utilizó una plataforma en la nube para el manejo tanto de la información como de los procesos que se pueden ejecutar en un servidor. Los servicios se implementaron en un proyecto de Google Cloud Platform.

Se tienen desplegados en dicho proyecto un servicio ftp el cual brinda el acceso a las imágenes para su procesamiento, se cuenta con una instancia de Compute Engine con una máquina escalable la cual tiene como OS ubuntu 18.

Sobre esta instancia se encuentra desplegada la base de datos referenciada en la sección anterior de este capítulo y a su vez se encuentran desplegados cada uno de los servicios de las capas que se manejan para el proceso de Multiview Learning.

Cada módulo del proyecto fue dockerizado y genera un entorno virtual con las dependencias necesarias para proveer los servicios de clasificación a demanda.

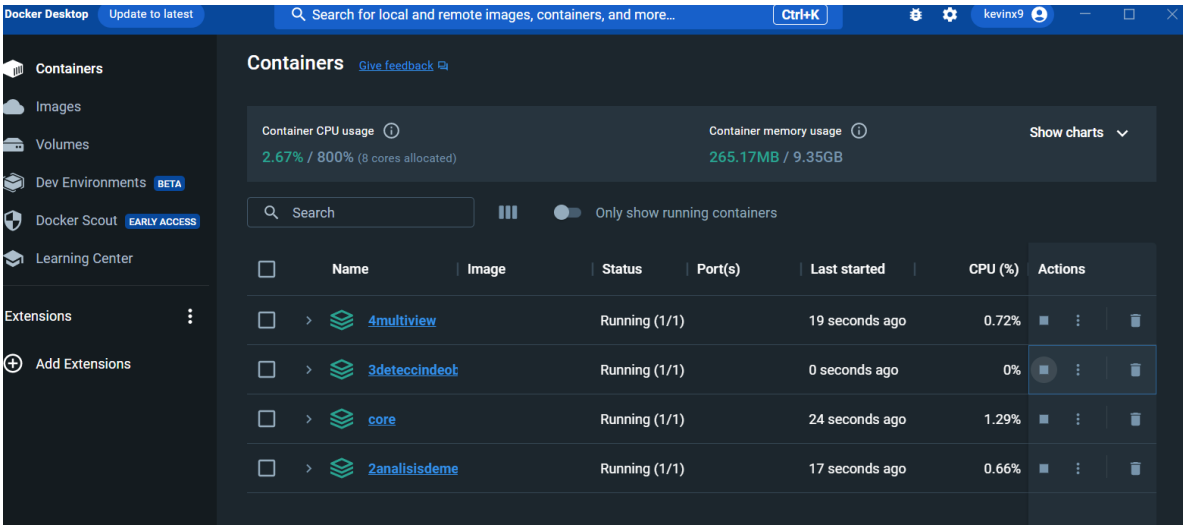

Ilustración 26: Visualización de los procesos de cada vista y de Multiview Dockerizados. Fuente: Autor

6.4 Implementación de la vista de Redes Neuronales Convolucionales

Para la implementación de esta primera vista la cual corresponde a la extracción de características visuales de alto nivel a partir del contenido de cada imagen. Para ello, se utilizó una red neuronal convolucional (CNN), esto a partir de la capacidad para capturar patrones visuales complejos como bordes, texturas, formas y estructuras jerárquicas.

Para esta implementación se establecieron dos posibles etiquetas (Violentas y No violentas) las cuales fueron la base ya que desde el punto de repositorio se tiene una colección de imágenes con sus respectivas subcarpetas.

Para el tema de pruebas se reorganizaron las muestras en una jerarquía estándar (data/train, data/val, data/test) con proporciones 70 / 15 / 15 %, asegurando aleatoriedad controlada mediante una semilla fija. Este paso fue automatizado por el script prepare_dataset.py, el cual facilita la carga directa con ImageDataGenerator.flow_from_directory, evita fugas de información entre particiones y mantiene trazabilidad sobre las etiquetas.

```python
import argparse, os, random, shutil, sys
from pathlib import Path

IMG_EXTS = {".jpg", ".jpeg", ".png", ".bmp", ".gif", ".tiff"}

def gather_images(class_dir: Path) -> list[Path]:

    return [p for p in class_dir.rglob("*") if p.suffix.lower() in IMG_EXTS]

def split_list(items, train_frac, val_frac):

    n = len(items)
    n_train = int(n * train_frac)
    n_val   = int(n * val_frac)
    train = items[:n_train]
    val   = items[n_train:n_train + n_val]
    test  = items[n_train + n_val:]
    return train, val, test

def distribute(files, dest_root: Path, subset: str, class_name: str, move=False):

    subset_dir = dest_root / subset / class_name
    subset_dir.mkdir(parents=True, exist_ok=True)

    for src in files:
```

Ilustración 27: Visualización del código de preparación del dataset. Fuente: Autor

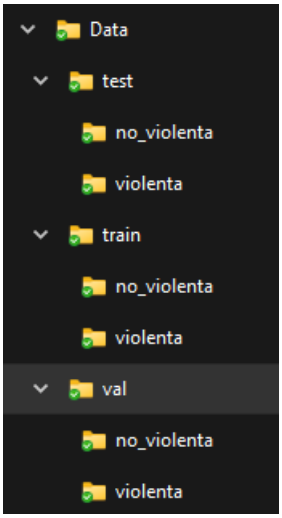

Ilustración 28: Vista de la jerarquía de carpetas para el manejo de las imágenes. Fuente: Autor

Durante este procesamiento se identifican si hay formatos de imágenes no soportados, esto con la finalidad de no afectar las pruebas con el dataset.

Posteriormente se creó un script en Python utilizando la librería de Tensorflow (TensorFlow, 2022) la cual es una librería que contiene modelos de deep learning entre los cuales se encuentran las Redes Neuronales Convolucionales CNN, esto siendo elegido debido a las diversas ventajas tales como:

- Capacidades de una fácil implementación con un lenguaje de programación sencillo como lo es Python
- Cuenta con funciones optimizadas para convoluciones en 2d que son esenciales para poder extraer las características de las imágenes como bordes, texturas y patrones.
- Cuenta con una abstracción de alto nivel con KERAS permitiendo la definición de una arquitectura CNN (con capas Conv2D, MaxPooling, BatchNormalization, entre otras)
- Cuenta con capacidad de escalabilidad, despliegue y robustes para entrenamiento, despliegue y visualización que se pueden explorar para mejorar el procesamiento.

Partiendo de esta definición de usar Tensorflow se creó un script de para la implementación de la red neuronal el cual se nombró como cnn_violencia.py

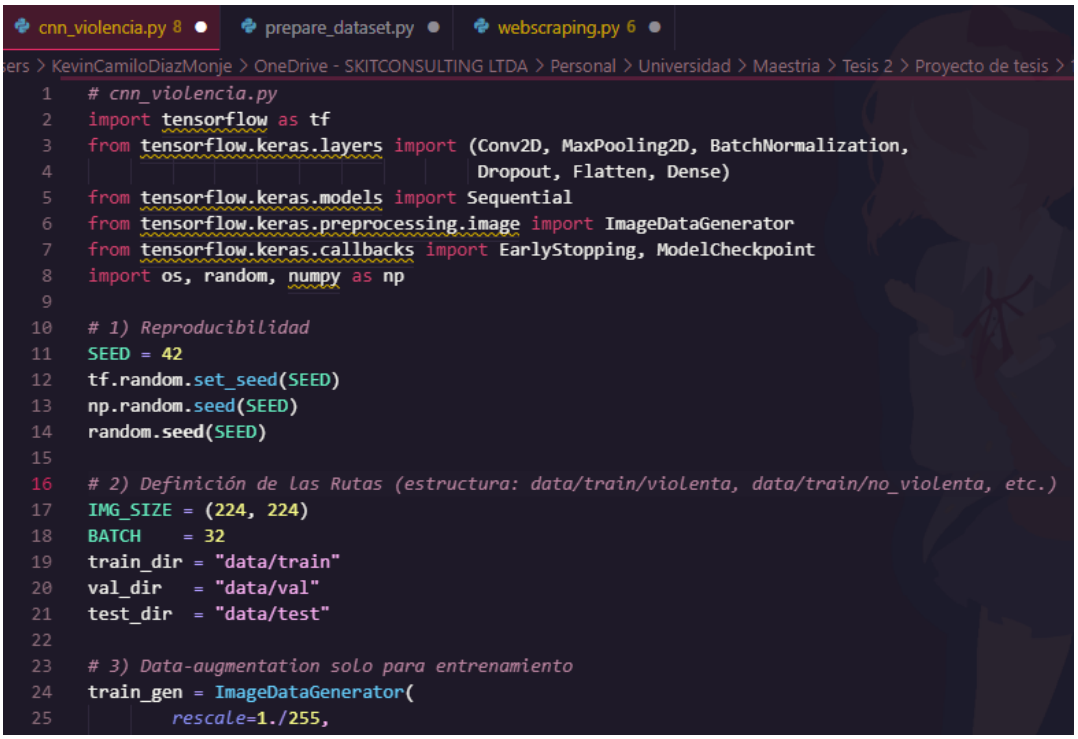

Ilustración 29: Visualización del código donde se estableció la red neuronal. Fuente: Autor

En este script se realizan varios pasos para el desarrollo y tratamiento de las imágenes en la red neuronal, para la cual se maneja una estructura de feed forward (es un tipo de red neuronal artificial que se utiliza en el aprendizaje automático y reconocimiento de patrones.) (gamco, s.f.) con capas convolucionales para la extracción de las características y hacer una clasificación binaria. A continuación, se da una breve explicación de los pasos realizados

1. Recolección de los datos: se establecen las rutas de los datasets
2. Se hace el preprocesamiento de la imagen que se está trabajando para ello se hace un redimensionamiento y una normalización de la imagen
3. Se aplicó la técnica de *data augmentation* (la cual implica una serie de rotaciones leves, desplazamientos, flips horizontales y variaciones de zoom y brillo) para mejorar la capacidad de generalización del modelo frente a variaciones de ángulo, escala y luminosidad que se puedan tener en las imágenes.
4. Se hace la definición del modelo CNN el cual está basado en tres bloques Conv2D → ReLU → BatchNorm → MaxPool → Flatten + Dense, seguida de capas densas con regularización Dropout
    a. La capa de convolución (Conv2D) lo que realiza es un deslizamiento de un conjunto de filtros 3×3 sobre la imagen. Cada filtro aprende a resaltar un patrón (bordes, texturas, armas, etc.). El resultado son mapas de activación que indican dónde aparece ese patrón.
    b. ReLU lo que realiza es una sustitución de los valores negativos que pueden aparecer por un 0 para así poder garantizar que las funciones no fallen al manejar temas como normalización.
    c. BarchNorm lo que realiza es una normalización de cada mini-batch para que así poder realizar una convergencia y estabilizar el entrenamiento
    d. Max Pooling hace una reducción en la resolución espacial, de tal manera que extrae el valor máximo de una ventana 2×2. Y así es posible mantener la característica dominante.

e. Flatten + Dense esta capa convierte los mapas de activaciones en vectores para posteriormente combinar las características aprendidas y producir una única probabilidad la cual está definida como sigmoid

f. Dropout realiza una desactivación de las neuronas de forma aleatoria en la fase de entrenamiento con la finalidad de disminuir un “Sobreajuste” que pueda tener la red neuronal.

5. Se estableció una función de pérdida enfocada en la parte binaria la cual fue binary_crossentropy, optimizada con “Adam” (es un algoritmo de optimización adaptativo que puede ajustar de forma óptima la tasa de aprendizaje de un modelo de aprendizaje automático en función de la magnitud del gradiente) (KrianVoleti, 2024),a una tasa de aprendizaje inicial de $1 \cdot 10^{-4}$

6. Se agregaron unos “*callbacks”* los cuales son “EarlyStopping” y “ModelCheckpoint” los cuales detienen el proceso de entrenamiento so se llega a un punto en donde ya no hay más posibles mejoras en la exactitud de validación y conservan la versión del modelo con mejor desempeño.

Ya con esta información se almacenaron los pesos en un archivo llamado “mejormodelo.h5” permitiendo que si se hace un refinamiento en el proceso de entrenamiento para casos específicos no se tenga una interferencia.

output.txt

```
Found 4015 images belonging to 2 classes.
Found 862 images belonging to 2 classes.
Found 862 images belonging to 2 classes.
/home/kevin/Downloads/red_neuronal_convolucional/venv/lib/python3.12/site-packages/keras/src/layers/convolutional/base_conv.py:113: UserWarning: Do not pass an `input_shape`/`input_dim`
  super().__init__(activity_regularizer=activity_regularizer, **kwargs)
2025-06-04 23:06:15.503915: E external/local_xla/xla/stream_executor/cuda/cuda_platform.cc:51] failed call to cuInit: INTERNAL: CUDA error: Failed call to cuInit: UNKNOWN ERROR (303)
Model: "sequential"
```

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 222, 222, 32)	896
batch_normalization (BatchNormalization)	(None, 222, 222, 32)	128
max_pooling2d (MaxPooling2D)	(None, 111, 111, 32)	0
conv2d_1 (Conv2D)	(None, 109, 109, 64)	18,496
batch_normalization_1 (BatchNormalization)	(None, 109, 109, 64)	256
max_pooling2d_1 (MaxPooling2D)	(None, 54, 54, 64)	0
conv2d_2 (Conv2D)	(None, 52, 52, 128)	73,856
batch_normalization_2 (BatchNormalization)	(None, 52, 52, 128)	512
max_pooling2d_2 (MaxPooling2D)	(None, 26, 26, 128)	0
flatten (Flatten)	(None, 86528)	0
dropout (Dropout)	(None, 86528)	0
dense (Dense)	(None, 128)	11,075,712
dropout_1 (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 1)	129

```
Total params: 11,169,985 (42.61 MB)
```

```
output.txt
/home/kevin/Downloads/red_neuronal_convolucional/venv/lib/python3.12/site-packages/keras/src/trainers/data_adapters/py_dataset_adapter.py:121: UserWarning: Your `PyDataset` class should ca
  self._warn_if_super_not_called()
Epoch 1/30
 64/126 ━━━━━━━━━━━━━━━━━━━━ 1:43 2s/step - accuracy: 0.5713 - loss: 1.4986/home/kevin/Downloads/red_neuronal_convolucional/venv/lib/python3.12/site-packages/PIL/Image.py:3442: DecompressionBom
  warnings.warn(
126/126 ━━━━━━━━━━━━━━━━━━━━ 0s 2s/step - accuracy: 0.5874 - loss: 1.2247/home/kevin/Downloads/red_neuronal_convolucional/venv/lib/python3.12/site-packages/keras/src/trainers/data_adapters/py_d
  self._warn_if_super_not_called()
/home/kevin/Downloads/red_neuronal_convolucional/venv/lib/python3.12/site-packages/PIL/Image.py:1043: UserWarning: Palette images with Transparency expressed in bytes should be converted t
  warnings.warn(
WARNING:absl:You are saving your model as an HDF5 file via `model.save()` or `keras.saving.save_model(model)`. This file format is considered legacy. We recommend using instead the native
126/126 ━━━━━━━━━━━━━━━━━━━━ 245s 2s/step - accuracy: 0.5876 - loss: 1.2218 - val_accuracy: 0.5615 - val_loss: 1.6001
Epoch 2/30
126/126 ━━━━━━━━━━━━━━━━━━━━ 251s 2s/step - accuracy: 0.7036 - loss: 0.5924 - val_accuracy: 0.4374 - val_loss: 5.4129
Epoch 3/30
126/126 ━━━━━━━━━━━━━━━━━━━━ 251s 2s/step - accuracy: 0.7205 - loss: 0.5461 - val_accuracy: 0.4490 - val_loss: 1.6569
Epoch 4/30
126/126 ━━━━━━━━━━━━━━━━━━━━ 0s 2s/step - accuracy: 0.7342 - loss: 0.5279WARNING:absl:You are saving your model as an HDF5 file via `model.save()` or `keras.saving.save_model(model)`. This file
126/126 ━━━━━━━━━━━━━━━━━━━━ 253s 2s/step - accuracy: 0.7342 - loss: 0.5279 - val_accuracy: 0.5882 - val_loss: 0.9174
Epoch 5/30
126/126 ━━━━━━━━━━━━━━━━━━━━ 0s 2s/step - accuracy: 0.7489 - loss: 0.5185WARNING:absl:You are saving your model as an HDF5 file via `model.save()` or `keras.saving.save_model(model)`. This file
126/126 ━━━━━━━━━━━━━━━━━━━━ 255s 2s/step - accuracy: 0.7489 - loss: 0.5184 - val_accuracy: 0.6705 - val_loss: 0.6598
Epoch 6/30
126/126 ━━━━━━━━━━━━━━━━━━━━ 256s 2s/step - accuracy: 0.7584 - loss: 0.4924 - val_accuracy: 0.5824 - val_loss: 1.0162
Epoch 7/30
126/126 ━━━━━━━━━━━━━━━━━━━━ 250s 2s/step - accuracy: 0.7594 - loss: 0.4806 - val_accuracy: 0.5545 - val_loss: 0.9833
Epoch 8/30
126/126 ━━━━━━━━━━━━━━━━━━━━ 0s 2s/step - accuracy: 0.7688 - loss: 0.4794WARNING:absl:You are saving your model as an HDF5 file via `model.save()` or `keras.saving.save_model(model)`. This file
126/126 ━━━━━━━━━━━━━━━━━━━━ 255s 2s/step - accuracy: 0.7688 - loss: 0.4794 - val_accuracy: 0.6810 - val_loss: 0.6521
Epoch 9/30
126/126 ━━━━━━━━━━━━━━━━━━━━ 257s 2s/step - accuracy: 0.7905 - loss: 0.4505 - val_accuracy: 0.6230 - val_loss: 0.8442
Epoch 10/30
126/126 ━━━━━━━━━━━━━━━━━━━━ 255s 2s/step - accuracy: 0.7611 - loss: 0.4796 - val_accuracy: 0.5974 - val_loss: 0.8703
Epoch 11/30
126/126 ━━━━━━━━━━━━━━━━━━━━ 0s 2s/step - accuracy: 0.7855 - loss: 0.4769WARNING:absl:You are saving your model as an HDF5 file via `model.save()` or `keras.saving.save_model(model)`. This file
126/126 ━━━━━━━━━━━━━━━━━━━━ 255s 2s/step - accuracy: 0.7855 - loss: 0.4768 - val_accuracy: 0.7042 - val_loss: 0.6772
Epoch 12/30
126/126 ━━━━━━━━━━━━━━━━━━━━ 256s 2s/step - accuracy: 0.7875 - loss: 0.4445 - val_accuracy: 0.5394 - val_loss: 1.2603
Epoch 13/30
126/126 ━━━━━━━━━━━━━━━━━━━━ 254s 2s/step - accuracy: 0.7936 - loss: 0.4365 - val_accuracy: 0.6462 - val_loss: 0.9038
27/27 ━━━━━━━━━━━━━━━━━━━━ 25s 898ms/step - accuracy: 0.7998 - loss: 0.4611
Exactitud final en test: 0.708
```

Ilustración 30: Resultado de la ejecución completa del proceso. Fuente: Autor

Es fundamental considerar la rigurosidad en la preparación del conjunto de datos (imágenes) y la transparencia en la configuración del modelo, ya que estos son factores clave para fortalecer la validez interna de los resultados. Con la configuración establecida de logró una exactitud del 70% lo cual es un margen aceptable para este tipo de clasificación debido a la diversidad de casos. Este el primer resultado obtenido para la primera de las capas planteadas en el esquema de *Multiview learning (MVL)*. Por lo cual se define el primer factor de entrada a el modelo de Multiview Learning.

$$R_c = [(id\ img_1 \quad value),\ \ (id\ img_2 \quad value),\ \ \ldots\ \ (id\ img_n \quad value)]$$

$$R_c = [(id\ img_1 \quad {}^{1}/_{0}),\ \ (id\ img_2 \quad {}^{1}/_{0}),\ \ \ldots\ \ (id\ img_n \quad {}^{1}/_{0})]$$

Donde:

$$id\ img_n = Id\ de\ la\ imagen\ evaluada$$

$$value = valor\ de\ la\ evaluación\ de\ la\ imagen\ (1\ es\ violenta, 0\ no\ es\ violenta)$$

Ecuación 1: Estructura del vector de resultados de la capa de Redes Neuronales convolucionales. Fuente: Autor

Teniendo este resultado cada imagen tendrá un valor entre 0 y 1 en donde el 0 significa que la imagen no es probable a que sea violenta mientras que el 1 representa que la imagen es una probabilidad de que si es violenta en el marco del caso de estudio.

Cabe recalcar que la valoración tomada para cada imagen se guarda en la base de datos para tener un referente del valor establecido al ejecutar esta primera vista sobre cada una de las imágenes.

6.5 Implementación de la vista de Análisis de metadatos

En este apartado se muestra la implementación de la segunda vista la cual corresponde al análisis de la información que se encuentra en los metadatos de cada una de las imágenes.

Esta vista está enfocada en poder realizar el procesamiento de un texto descriptivo como lo es el nombre y/o descripción de la imagen siendo este nuestro metadato sobre el cual se trabajará, una vez obtenido este metadato el objetivo es poder calcular un valor numérico que logre representar la complejidad contextual de este campo enfocada a determinar si se encuentran posibles palabras que representen un aspecto violento dentro del contexto del caso de estudio (conflicto armado en Colombia)

Para realizar el respectivo análisis se establecieron dos diccionarios de información los cuales se realizaron basados en la información que es relevante para el contexto, de tal forma que si se quiere expandir el uso de esta vista a otro contexto se debe generar un diccionario de pesos y de conectores para su uso adecuado al contexto a trabajar, a continuación, se explica cada uno de los Diccionarios:

- Diccionario de pesos: el cual contiene por cada renglón un arreglo de palabra y peso, en donde se define una palabra que esté enmarcada como violenta para el contexto y un peso de cuan violenta puede ser.

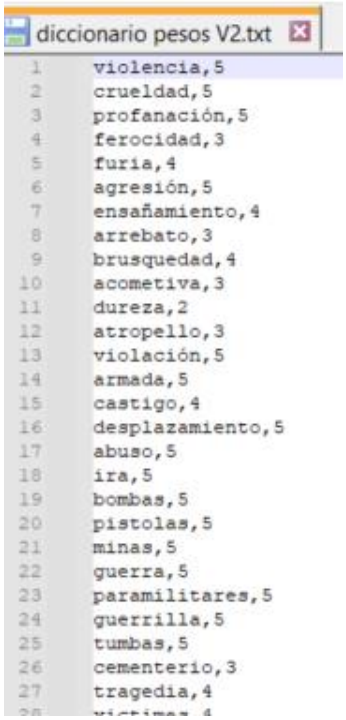

Ilustración 31: Visualización del diccionario de pesos. Fuente: Autor

- Diccionario de conectores: Este contiene palabras que dentro de la semántica funcionan como conectores y por si solos no tienen ninguna implicación de inferencia hacia hacer alusión a la violencia en el contexto del conflicto armado

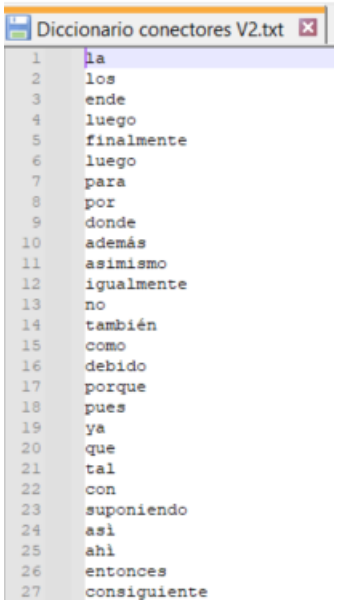

Ilustración 32: Visualización del diccionario de conectores. Fuente: Autor

Continuando con el proceso se detallan los siguientes pasos

1. Conexión a la base de datos
2. Extracción del metadato= 9 por cada una de las imágenes
3. Se eliminan signos y tildes de cada palabra con la finalidad de normalizar la cadena de texto para su análisis por cada palabra
4. Se establecen ciertos parámetros para cada cadena
    a. pt: total de palabras de la imagen.
    b. pd: cantidad de palabras descartadas.
    c. pu: palabras que son usadas.
    d. pe: palabras encontradas en el diccionario de pesos.
    e. p: suma de pesos de las palabras encontradas.
5. Para las palabras que se encontraron en el diccionario de pesos se obtiene el valor del peso que se
6. Se establece la fórmula para determinar el resultado del análisis

$$Resultado = \left| \frac{\left(\frac{p}{pe+1}\right) - (pu - pe)}{(pt - pd) + 1} \right|$$

Ecuación 2: Cálculo del resultado de análisis de metadato. Fuente: Autor

Este cálculo lo que busca es poder obtener todos los pesos encontrados y dividirlos entre la cantidad de las palabras más 1, para posteriormente restar la diferencia entre las palabras usadas para el análisis y las que, si obtuvieron un peso, todo esto entre la diferencia entre el total de las palabras y las que fueron descartadas (las que están en el diccionario de conectores) + 1, para al final normalizar el valor con un valor absoluto.

A continuación, se colocan algunos ejemplos del cálculo del resultado del análisis

Metadato	Peregrinación en el Monumento a las víctimas de Trujillo, 2008	**Cantidad**
	Pt = Total de palabras de la imagen	10
	Pd= Total de palabras descartadas	6
	Pu= Total de palabras usadas en el análisis	4
	Pe = Total de palabras encontradas como violentas	1
	Resultado	0,5

Metadato	Memorias del éxodo en la guerra	**Cantidad**
	Pt = Total de palabras de la imagen	6
	Pd= Total de palabras descartadas	3
	Pu= Total de palabras usadas en el análisis	3
	Pe = Total de palabras encontradas como violentas	2
	Resultado	0,83333333

Metadato	A los 21 años, Gloria Aristizábal fue desaparecida por los paramilitares. Su esposo, José Alberto Franco, había sido asesinado un año antes. Los 4 hijos de este matrimonio, quedaron a cargo de su abuela Rosalba Aristizábal	**Cantidad**
	Pt = Total de palabras de la imagen	36
	Pd= Total de palabras descartadas	14
	Pu= Total de palabras usadas en el análisis	22
	Pe = Total de palabras encontradas como violentas	3
	Resultado	0,66666667

Tabla 1: Ejemplos de la forma en la cual se analiza la descripción de una imagen. Fuente: Autor

Una vez establecido el valor de cada imagen este tendrá un valor entre 0 y 1 en donde el 0 significa que la meta data de la imagen que se está trabajando no contiene ningún contenido que haga que sea violenta mientras que el 1 representa que la meta data de la imagen contiene palabras que si hacen alusión a la violencia en el marco del caso de estudio.

Al igual que en la capa anterior se define el factor de entrada a el modelo de Multiview learning

$$R_t = \left[\left(id\ img_1 \quad R_{(Metadata)_1}\right), \quad \left(id\ img_2 \quad R_{(Metadata)_2}\right), \quad \ldots \quad \left(id\ img_n \quad R_{(Metadata)_n}\right)\right]$$

$$R_t = [(id\ img_1 \quad 0{,}5), \quad (id\ img_2 \quad 0{,}83), \quad \ldots \quad (id\ img_n \quad 0{,}67)]$$

Donde:

$$id\ img_n = Id\ de\ la\ imagen\ evaluada$$

$$R_{(Metadata)_n} = Resultado\ de\ la\ evaluaciónd\ e\ metadata\ (1\ es\ mas\ violenta, 0\ es\ menos\ violenta$$

Ecuación 3: Estructura del vector de resultados de la capa de análisis de metadatos. Fuente: Autor

Cabe recalcar que la valoración tomada para cada imagen se guarda en la base de datos para tener un referente del valor establecido al ejecutar esta segunda vista sobre la meta data de cada una de las imágenes.

6.6 Implementación de la vista de Detección de Objetos

En esta sección se muestra la implementación de la tercera vista la cual corresponde a la detección de objetos que se encuentra en cada una de las imágenes.

Esta vista está enfocada en poder determinar si alguno de los objetos que se establecieron como “Violentos” dentro del contexto de caso de estudio del conflicto armado en Colombia se encuentran en la imagen que se está procesando, esto con la finalidad de lograr identificar que contenido como tal tiene la imagen y así clasificar la imagen a partir de un valor que estará determinado por la cantidad de objetos detectados en la imagen.

Para esta vista se utilizó YOLOv4 (You Only Look Once versión 4), un modelo ampliamente utilizado en visión por computador por su alta precisión a la hora de establecer objetos. Para este caso se apoya en la biblioteca Darknet y en OpenCV para cargar, procesar imágenes y visualizar resultados.

6.6.1 Estructuración y entrenamiento

Primeramente, se inició por estructura los directorios que se iban a manejar para localizar los datos

```
darknet/
 ├── build/
 ├── data/
|   ├── obj/              <-- Carpeta que contiene todas las imágenes y sus .txt
|   |   ├── Ejemplo de imagen 1.jpg
|   |   ├── Ejemplo de imagen 1.txt
|   |   └── ...
|   ├── obj.names         <-- Archivo con los nombres de las clases
|   ├── obj.data          <-- Archivo de configuración de datos
|   ├── train.txt         <-- Carpeta de imágenes de entrenamiento
|   └── valid.txt         <-- Carpeta de imágenes de validación
 ├── cfg/
|   └── yolov4-custom.cfg   <-- Archivo de configuración de la arquitectura
└── backup/               <-- Directorio donde se guardan los pesos del modelo
```

Se continuó con una división de datos en donde se estableció un 70% para entrenamiento (train) y un 30 por ciento para validación (valid), los archivos train.txt y valid.txt contienen con las rutas relativas a cada una de las imágenes correspondientes a estos conjuntos.

Para configurar la arquitectura del modelo se tiene el archivo obj.names quien tiene la definición de algunos objetos que serán los que se van a trabajar donde cada nombre de

clase se escribe en una nueva línea. El número de línea corresponde al class_id utilizado en los archivos de anotación

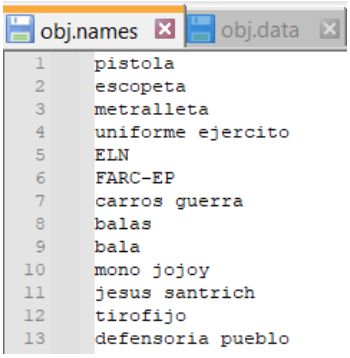

Ilustración 33: Nombres de los objetos a trabajar. Fuente: Autor

Se cuenta con un archivo obj.data este archivo le indica a Darknet dónde encontrar todos los recursos necesarios para el entrenamiento, desde la clase las rutas etc.

- # Contenido de obj.data
- classes = [Número de clases]
- train = data/train.txt
- valid = data/valid.txt
- names = data/obj.names

Se cuenta con el archivo yolov4-custom.cfg, el cual es el más importante porque tiene toda la definición de la arquitectura completa de la red neuronal

Aquí se tienen algunos parámetros como:

1. **Parámetros de Entrenamiento:**
   - batch=64: es el número de imágenes procesadas en cada lote.
   - subdivisions=16: Divide el lote en “mini-lotes” para ajustarse a la memoria de la GPU.
   - width=416 y height=416: es la resolución de entrada de la red. Se eligió un múltiplo de 32 para compatibilidad con la arquitectura.
2. **Parámetros Específicos del Problema:**
   - max_batches = [Número de clases] * 2000: es el número total de iteraciones de entrenamiento. La regla general es 2000 iteraciones por clase, pero no menos de 6000.
   - steps = [80% de max_batches], [90% de max_batches]: esto define en qué

iteraciones la tasa de aprendizaje se reduce.

- classes = [Número de clases]: Se actualizó el número de clases en las **tres capas [yolo]** presentes en el archivo.
- filters = ([Número de clases] + 5) * 3: Justo antes de cada una de las tres capas [yolo], hay una capa [convolutional]. El parámetro filters de estas tres capas se recalculó con esta fórmula.

Para la parte de entrenamiento se debe ejecutar desde la terminal en el directorio raíz de Darknet

```
./darknet detector train data/obj.data cfg/yolov4-custom.cfg
yolov4.conv.137                -dont_show                -map
```

Durante el entrenamiento, se revisaron las siguientes métricas clave:

- **Average Loss:** El error promedio del modelo.
  Se revisa que la métrica disminuya de forma constante y se estabilizara en un valor bajo (idealmente por debajo de 1.0).
- **mAP@0.5:** La métrica principal de rendimiento.
  Indica la precisión promedio del modelo con un umbral de IoU (Intersection over Union) de 0.5. para que de esta manera se pueda maximizar este valor.

El entrenamiento se detuvo en el momento en que la average loss dejó de presentar una disminución significativa y el mAP en el conjunto de validación se estabilizó, lo que indicó que el modelo había alcanzado la convergencia.

Con estos valores, Darknet guarda automáticamente los pesos del modelo en el directorio backup/ cada 1000 iteraciones, además de generar el archivo yolov4-custom_best.weights, correspondiente al modelo con el valor de mAP más alto registrado hasta ese punto.

6.6.2 Validación

Conforme a los resultados obtenidos en la fase de entrenamiento y con la estructura planteada anteriormente se procede a la ejecución de la validación, para lo cual el módulo realiza los siguientes pasos:

1. Se cargan las diversas librerías que se mencionaron anteriormente
2. Se deben tener una preclasificación de las imágenes (información generada en la fase de entrenamiento) para lo cual se generó un archivo de configuración y de mejores pesos de forma similar como se generó en la primera capa esto con la finalidad de ya tener entrenado el sistema con los objetos ya clasificados, para ello

se le asigna una carpeta a cada uno de los objetos para tener las referencias de cada uno de los objetos

3. Ya dentro del script principal se utiliza la función detect_objects_in_image la cual toma una ruta de imagen, una red neuronal cargada, nombres de clases y colores de clases, y realiza el siguiente proceso por cada imagen:
    a. Comienza con la obtención del tamaño requerido por la red.
    b. Lee la imagen con OpenCV.
    c. Posteriormente convierte la imagen a formato RGB y la redimensiona.
    d. Copia la información en bytes de imagen a una estructura de Darknet.
    e. Sigue con la detección utilizando la función detect_image.
    f. Se libera memoria y dibuja los cuadros delimitadores (también llamados draw_boxes) en la imagen redimensionada.
    g. Se Restaura el tamaño original de la imagen para visualización final.
    h. Devuelve la información de detecciones procesada por get_info.
4. Se usa la función get_info la cual obtiene el nombre de la imagen y devuelve una lista de objetos detectados (nombre, confianza y coordenadas del recuadro).
5. Posteriormente se hace un conteo de los objetos por clase usando la función how_many_objects_in_image.
6. La función how_many_objects_in_image hace el conteo de cuántos objetos fueron detectados por clase en la imagen.

Como resultado de aplicar este modelo se obtiene una salida en donde se muestra cada uno de los objetos encontrados, con su pertinencia y los 4 puntos que conforman el recuadro, y también que cantidad de objetos se encontraron

Resultado 1.coffee ● webscraping.py 6 ● modelo_final_yolo_v4.py 9+

C: > Users > KevinCamiloDiazMonje > OneDrive - SKITCONSULTING LTDA > Personal > Universidad > Maestria > Tesis 2 > Proyecto de tesis > 3. Det


```
{'image_name': 'imagen_0', 'detected_objects':
[{'object': 'tirofijo',
  'confidence': '97.61',
  'square_points': (125.52667236328125, 120.56813049316406, 93.54959106445312, 134.70387268066406)},
{'object': 'mono jojoy',
 'confidence': '99.26',
 'square_points': (303.56671142578125, 82.82205200195312, 93.95613861083984, 124.96743774414062)}],
 'quantity_objects':
    {'pistola': 0,
    'escopeta': 0,
    'metralleta': 0,
    'uniforme ejercito': 0,
    'ELN': 0,
    'FARC-EP': 0,
    'carros guerra': 0,
    'balas': 0,
    'bala': 0,
    'mono jojoy': 1,
    'jesus santrich': 0,
    'tirofijo': 1,
    'defensoria pueblo': 0,
    'cruz roja': 0,
    'paloma paz': 0}}
```


Ilustración 34: Resultado de un proceso de detección de objetos sobre una imagen. Fuente: Autor

Ilustración 35: Imagen de referencia para el ejemplo de detección de objetos. Fuente: Autor

Posteriormente cuando ya se obtiene por cada imagen un resultado tal como se presentaba anteriormente se hace la sumatoria de la cantidad de veces que se encontró un objeto y posteriormente la sumatoria de todos esos objetos para así poder dar un valor de la identificación de objetos

$$R_{o_{id\ img_n}} = \sum_{i=1}^{id\ img_n} \sum_{j=1}^{Obj_m} value$$

Donde:

$$id\ img_n = Id\ de\ la\ imagen\ evaluada$$

$$Obj_m = id\ del\ objeto\ que\ se\ está\ buscando$$

$$value = valor\ de\ la\ evaluación\ del\ objeto\ en\ la\ imagen\ (1\ está, 0\ no\ está)$$

Ecuación 4: Cálculo del resultado de identificación de objetos. Fuente: Autor

Al igual que en la capa anterior se define el factor de entrada a el modelo de Multiview learning

$$R_0 = [(id\ img_1 \quad R_{o_1}), \quad (id\ img_2 \quad R_{o_2}), \quad \ldots \quad (id\ img_n \quad R_{o_n})]$$

Ecuación 5: Estructura del vector de resultados de la capa de Detección de Objetos. Fuente: Autor

Cabe recalcar que la valoración tomada para cada imagen se guarda en la base de datos para tener un referente del valor establecido al ejecutar esta segunda vista sobre la meta data de cada una de las imágenes.

6.7 Implementación del módulo de Multiview Learning

Una vez se tengan las respectivas clasificaciones de las tres capas principales anteriormente explicadas se procede a establecer la forma en la cual se va a realizar la condensación de la información por cada imagen, Aquí es donde el Multiview se hace presente a través de una red neuronal más sencilla de tipo perceptrón el cual a partir de los resultados numéricos extraídas de las imágenes (representadas por tres atributos: Rc, Rt y Ro), el algoritmo es entrenado para distinguir entre imágenes violentas y no violentas. Permitiendo que el funcionamiento del perceptrón se use como clasificador binario.

Para dar un contexto del funcionamiento se establece la forma en la cual se manejan los factores de cada una de las vistas y se le asigna un peso a cada una de ellas (esto basado en el entrenamiento del perceptrón)

$$R_m = (R_c * P_a) + (R_t * P_b) + R_o * P_c$$

Donde:

$$R_c = Resultado\ del\ modulo\ de\ redes\ convolucionales$$

$$R_t = Resultado\ del\ modulo\ de\ análisis\ de\ metadatos$$

$$R_o = Resultado\ del\ modulo\ de\ Detección\ de\ objetos$$

$$R_m = Resultado\ del\ modulo\ de\ Multiview\ Learning$$

$$P_a, P_b, P_c = Pesos\ asignados\ a\ casa\ vista$$

Ecuación 6: Cálculo para el resultado del Multiview learning. Fuente: Autor

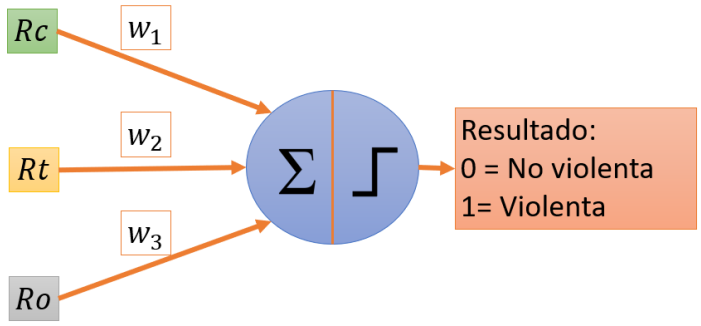

Ilustración 36: Diagrama del funcionamiento del perceptrón. Fuente: Autor

Para este proceso de clasificación el perceptrón obtiene los datos de un conjunto de imágenes de entrenamiento (para el ejemplo son 10 imágenes) con los posibles resultados de su clasificación (esto debido a que es un modelo supervisado).

Este proceso de entrenamiento realiza los siguientes pasos:

1. Se realiza una matriz de 3 * n donde n es el número de imágenes y el 3 hace referencia a los 3 valores que se tienen de resultado de las vistas previas Rc, Rt y Ro
2. Se realiza un arreglo de 1 por n en donde n es la misma cantidad de imágenes y el 1 hace referencia a los posibles valores de resultado que tiene su correspondiente imagen, en donde 1 es una imagen que se definió como “Violenta” y 0 si no se definió como “No violenta”
3. Se implementa una función de activación tipo escalón, que define la salida del perceptrón: 1 si la combinación lineal de entradas y pesos supera el umbral (b), 0 en caso contrario.
4. Se inicializan los pesos aleatoriamente y se define un umbral (b). Se fija una tasa de aprendizaje (0.01) y el número de épocas de entrenamiento (100). Tener en cuenta que todos estos factores pueden variar.
5. Se establece la parte del entrenamiento del perceptrón en donde
    a. Se calcula la predicción con la función de activación.
    b. Se mide el error (diferencia entre la clase real y la predicción).
    c. Se actualizan los pesos y el sesgo proporcionalmente al error.
6. El Resultado de cada uno de los pesos que se definieron se almacenan para hacer uso de ellos al momento de hacer la validación del modelo para este caso los pesos definidos fueron

```python
#Importaación de librerias
import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import axes3d

# Datos de 10 imagenes -> [Rc, Rt, Ro]

imagenes = np.array([[0.9, 0.0, 0.1], [0.4, 0.1, 0.5],
                     [0.3, 0.2, 0.3], [0.4, 0.1, 0.2],
                     [0.5, 0.2, 0.3], [0.5, 0.8, 0.9],
                     [1.0, 0.8, 0.7], [0.7, 0.7, 0.7],
                     [0.8, 0.8, 0.8], [0.9, 0.9, 0.9]])

#Se asigna x,y & z --> (Rc, Rt, Ro)
x=np.array(imagenes[:,0])
y=np.array(imagenes[:,1])
z=np.array(imagenes[:,2])

# 1 : Violenta    0 : No violenta
clases = np.array([0, 0, 0, 0, 0, 1, 1, 1, 1, 1])


#-----------------------------------------------
# Gráfica de dispersión (Rc, Rt, Ro)
fig = plt.figure(figsize=(12, 12))
ax = fig.add_subplot(projection='3d')


# Creating color map
my_cmap = plt.get_cmap('hsv')

# Creating plot
sctt = ax.scatter3D(x, y, z, alpha = 0.8, c = (x + y + z), cmap = my_cmap, marker ='^')
```

Ilustración 37: Vista del código de entrenamiento para el perceptrón. Fuente: Autor

Para hacerlo un poco más gráfico se hizo una representación del comportamiento de los valores de cada una de las capas en una gráfica 3D, en donde se puede evidenciar que las imágenes con mayore valores son las que han tenido clasificaciones en cada una de sus vistas como “Violentas”

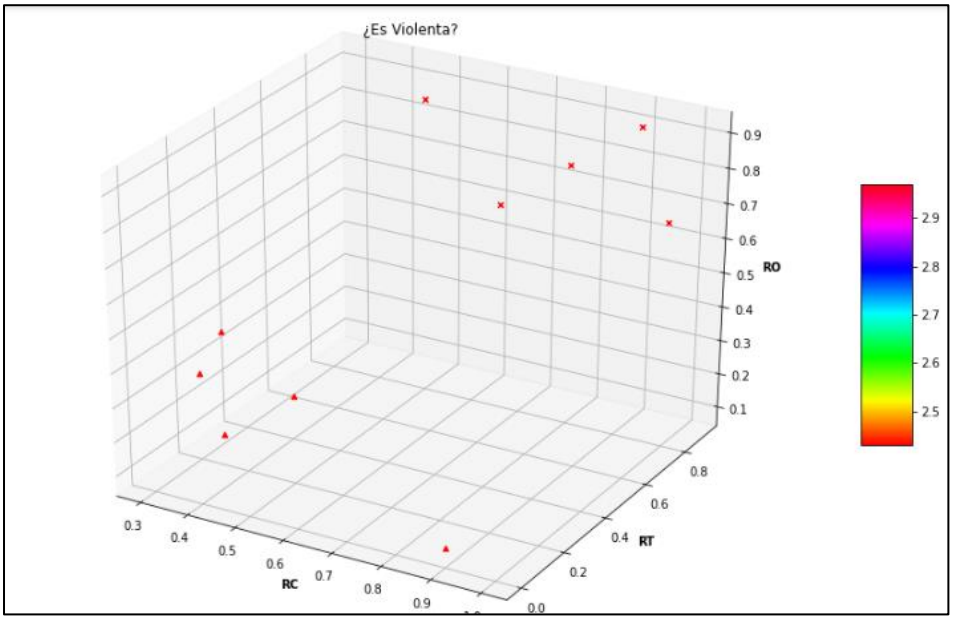

Ilustración 38: Diagramación de los resultados de cada una de las vistas. Fuente: Autor

6.8 Implementación del sitio web de ejecución

Esta es una breve aplicación web la cual busca poder hacer más amigable el uso de cada una de las capad del modelo de manera en que no se tenga que adentrar como tal en un entorno de desarrollo sino hacer uso de esta página para la ejecución de los procesos de cada una de las vistas

Ilustración 39: Visualización de la aplicación web. Fuente: Autor

En esta aplicación se cuenta con las 4 opciones para ejecutar cada vista, y también la del módulo de Multiview, y a su vez se tiene tienen dos opciones para poder ejecutar una cantidad de imágenes seleccionadas o para poder procesar todas las imágenes bajo una misma capa

Ilustración 40: Vista de la sección para ejecución de procesos por vista del modelo. . Fuente: Autor

Al hacer la ejecución de alguna de las dos formas esta información será almacenada en la base de datos para su posterior utilización.

6.9 Implementación de la vista de retroalimentación

Esta es una vista que se agrega a partir de la presentación de las imágenes en el sistema de visualización, para el caso de estudio se integró un sistema de visualización de trabajado en el trabajo de Tesis de Maestría "*DISEÑO DE UN METAMODELO DE SOFTWARE PARA EL DESARROLLO DE APLICACIONES INMERSIVAS*" (Sánchez Cruz, 2025), el cual está enfocado en la realidad aumentada, el prototipo que se desarrolló buscó realizar un libro en donde se puede mostrar las imágenes que se clasificaron, para ello la aplicación crea recorridos con imágenes ya clasificadas,

En la aplicación los usuarios pueden puntuar cada ítem de 1 a 5 para saber cómo le pareció la imagen, esto teniendo en cuenta que es una aplicación de Colombia histórica, dichos valores se guardarán para tenerse en cuenta como la nueva vista.

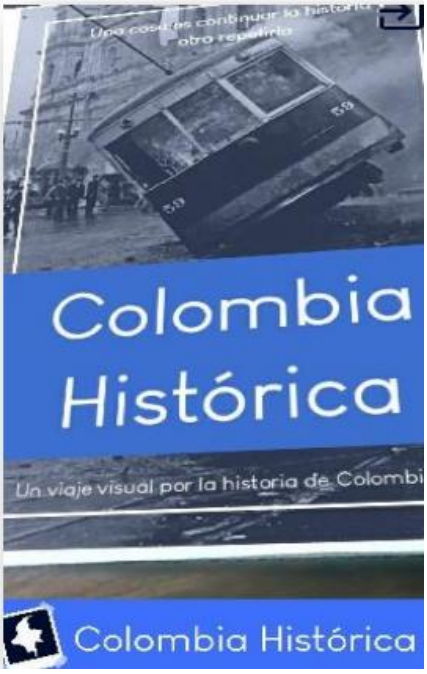

Ilustración 41: Imagen de presentación de la aplicación del sistema de visualización. Fuente: "DISEÑO DE UN METAMODELO DE SOFTWARE PARA EL DESARROLLO DE APLICACIONES INMERSIVAS" (Sánchez Cruz, 2025)

Ilustración 42: Imagen de la aplicación en la parte de calificación de la imagen. Fuente: "DISEÑO DE UN METAMODELO DE SOFTWARE PARA EL DESARROLLO DE APLICACIONES INMERSIVAS" (Sánchez Cruz, 2025)

Para poder hacer la conexión entre la aplicación y la base de datos se establecieron unos servicios para obtener las imágenes con un nivel de violencia y hacer que se presenten las imágenes que son acordes a la temática del sistema de visualización

```
def GetRandomImagenes(self, cantidad):
    return  [ imagen.json() for imagen in Imagen.query.order_by(func.random()).limit(cantidad).all()]

def GetRandomImagenesNivelViolencia(self, cantidad, nivelViolencia):
    imagenes = Imagen.query.filter_by(NivelViolencia=nivelViolencia).count()
    if(imagenes > 0):
        return  [ imagen.json() for imagen in Imagen.query.filter_by(NivelViolencia=nivelViolencia).order_by(func.random()).limit(cantidad).
    else:
        print("No existen imágenes con nivel de violencia "+nivelViolencia)
        return  [ imagen.json() for imagen in Imagen.query.order_by(func.random()).limit(cantidad).all()]

def GetRandomImagenesEsViolenta(self, violenta,cantidad):
    return  [ imagen.json() for imagen in Imagen.query.filter_by(EsViolenta=violenta).order_by(func.random()).limit(cantidad).all()]

def GetRandomImagenesEsViolentaNivelViolencia(self, violenta,cantidad, nivelViolencia):
    imagenes = Imagen.query.filter_by(NivelViolencia=nivelViolencia).count()
    if(imagenes > 0):
        return  [ imagen.json() for imagen in Imagen.query.filter_by(NivelViolencia=nivelViolencia,EsViolenta=violenta).order_by(func.rando
    else:
        print("No existen imágenes con nivel de violencia "+nivelViolencia)
        return  [ imagen.json() for imagen in Imagen.query.order_by(func.random()).limit(cantidad).all()]
```

Ilustración 43: Visualización del código de conexión para el consumo de las imágenes por parte del sistema de visualización. Fuente: Autor

también se hizo la función para el guardado de la calificación de la Imagen

```
def CalificarImagen (self,idRecorrido, idImagen, calificacion):

    imagenRecorrido = ImagenRecorrido.query.filter_by(IdRecorrido =idRecorrido,IdImagen = idImagen).first()

    if imagenRecorrido is None:
        return imagenRecorrido

    if calificacion is not None:
        imagenRecorrido.Calificacion = calificacion

    try:
        db.session.commit()
        return imagenRecorrido
    except Exception as e:
            print("ERROR: "+str(e))
            return False
```

Ilustración 44: Visualización del código para envió de la Calificación. Fuente: Autor

Ya conforme a cómo se integra con el proceso de Multiview se tiene en cuenta que se maneja como un refuerzo sobre el valor ya clasificado, siendo un Cálculo del promedio de las respuestas sobre cada imagen para que posteriormente se ingrese como un nuevo valor en el perceptrón para si respectiva asignación de pesos y posterior recalculo de la clasificación final por parte del Multiview

$$R_{m2} = (R_c * P_a) + (R_t * P_b) + (R_o * P_c) + (R_r * P_d)$$

Donde:

$$R_c = Resultado\ del\ modulo\ de\ redes\ convolucionales$$

$$R_t = Resultado\ del\ modulo\ de\ análisis\ de\ metadatos$$

$$R_o = Resultado\ del\ modulo\ de\ Detección\ de\ objetos$$

$$R_{m2} = Resultado\ del\ modulo\ de\ Multiview\ Learning\ con\ la\ cuarta\ vista$$

$$R_r = promedio\ de\ los\ resultados\ de\ la\ calificaciones\ registradas\ para\ una\ imagen$$

$$P_a, P_b, P_c, P_d = Pesos\ asignados\ a\ casa\ vista$$

Ecuación 7: Cálculo para el resultado del Multiview learning con la cuarta vista. Fuente: Autor

Para lo cual se aplica el mismo proceso, pero en vez de realizar un proceso con una matriz de 3*n se realiza un proceso de 4*n para establecer los nuevos pesos y así tener en cuenta la nueva capa y asignarle un peso.

Capítulo 7

Métricas y análisis de resultados

«No importa que nunca encuentre el corazón lo que ha buscado de verdad, no importa el tiempo que ya es muy corto en las ansias largas de vivir, cualquier minuto de placer será sentido en realidad.
Si lleno el alma, si lleno el alma de eternidad»

Autor: Diomedes Díaz

7 Métricas y análisis de resultados

7.1 Métricas

En el presente capitulo se busca presentar los diversos conceptos que se usaron para validar la efectividad del modelo de clasificación basado en Multiview Learning, para realizar esta tara se ha utilizado un conjunto de métricas que permiten validar el desempeño desde diversas perspectivas para así tener un amplio espectro de los resultados.

El conjunto de métricas que se manejaron se enfocó en la precisión, en estándares como lo son el F1-Score, Recall y hasta una expectativa por parte del usuario. Esto también porque a modo de comparación se tomaron las métricas manejadas en el artículo "*Multi-view learning for visual violence recognition with maximum entropy discrimination and deep features discrimination and deep features*" de (Sun, Liu, & Mao, 2018), con la finalidad de manejar métricas iguales para la validación del modelo.

A continuación, se presentan cada una de ellas cómo se aplicó para este proyecto.

7.1.1 Métrica de precisión

La métrica de precisión también conocida como Acurrancy o de exactitud, lo que busca es medir que porcentaje del total de las clasificaciones que hizo el modelo fueron correctas es decir cuantas veces el modelo al ejecutarse acertó en la clasificación

Este modelo está definido de la siguiente manera:

$$Precisión = \frac{Número\ de\ Clasificaciones\ Correctas}{Total\ de\ imágenes\ Clasificadas}$$

Ecuación 8: Formula de la métrica de precisión. Fuente: "Multi-view learning for visual violence recognition with maximum entropy discrimination and deep features discrimination and deep features" de (Sun, Liu, & Mao, 2018)

Para lograr hacer uso de esta métrica se tomaron los resultados obtenidos en la clasificación realizada por el Módulo de Multiview learning, para lo cual se logró obtener los siguientes datos:

Etiqueta de la clasificación	Resultado Correcto	Resultado No exacto	Total general
0	921	105	1026
1	2940	726	3666
Total general	**3861**	**831**	**4692**

Tabla 2: Tabla de resultados de clasificaciones Correctas. Fuente: Autor

Con base en estos datos se puede hacer el uso de la fórmula de precisión

$$Precisión = \frac{3861}{4692} = 0.8228$$

Ecuación 9: Resultado de la métrica de precisión. Fuente: Autor

Con lo anterior se muestra que el proceso de Multiview bajo estos parámetros cumple con un 82% de clasificaciones correctas conforme a la expectativa generada en la preclasificación realizada dentro del repositorio.

Nota: Cabe resaltar que de todo el conjunto de imágenes solo se lograron tener 4692 con las clasificaciones realizadas.

7.1.2 Métrica F1-Score

La métrica F1-Score es una métrica que combina la Precisión y un valor de Recall o sensibilidad, enfocándose en ver que tan preciso fue el modelo para una clase y también entendiendo cuales, si eran de esa clase, para ello esta métrica es definida como

$$F1\ score = 2 * \frac{Precisión * Recall}{Presición + Recall}$$

Ecuación 10: Fórmula de la métrica de F1-score. Fuente: "Multi-view learning for visual violence recognition with maximum entropy discrimination and deep features discrimination and deep features" de (Sun, Liu, & Mao, 2018)

Donde:

$$Precisión = \frac{Imágenes\ que\ cumplen\ la\ condición}{Imágenes\ que\ el\ modelo\ dijo\ que\ cumplia\ la\ condición}$$

Ecuación 11: Fórmula de la métrica de precisión para la métrica F1-score. Fuente: Autor

$$Recall = \frac{Imágenes\ que\ cumplen\ la\ condición}{Total\ de\ imágenes\ que\ Si\ cumplian\ la\ condición}$$

Ecuación 12: Fórmula de la métrica de Recall para la métrica F1-score. Fuente: Autor

Con esta base se comienzan a obtener cada una de las variables que se requieren para esta métrica:

1. Imágenes que el modelo dijo que, si cumplían con la condición, en este caso cuales el modelo dijo que 2940 imágenes eran correctas para la etiqueta 1
2. Imágenes que cumplen la condición, en este caso de esas 2940 se tenían que 2531 si cumplían la etiqueta 1
3. Total, de las imágenes que cumplen la condición; para el caso fueron 3666 imágenes que se tenían la etiqueta 1

Con lo anterior se resuelven las ecuaciones necesarias para obtener las variables de la métrica F1-score

$$Precisión = \frac{2531}{2940} = 0.86 = 86\%$$

Ecuación 13: Resultado de la fórmula de la métrica de precisión para la métrica F1-score. Fuente: Autor

$$Recall = \frac{2531}{3666} = 0.69 = 69\%$$

Ecuación 14: Fórmula de la métrica de Recall para la métrica F1-score. Fuente: Autor

Ya con estos dos valores se aplican en la fórmula de F1-score

$$F1\ score = 2 * \frac{0.86 * 0.69}{0.86 + 0.69} = 0.7662 = 77\%$$

Ecuación 15: Fórmula de la métrica de F1-score. Fuente: Autor

Con lo anterior se muestra que el proceso de Multiview bajo esta métrica cumple con un 77% en la precisión de las clasificaciones esperadas conforme a la expectativa generada en la preclasificación realizada dentro del repositorio.

7.1.3 Métrica de satisfacción del usuario

Esta métrica está basada en la retroalimentación realizada por una serie de usuarios que probaron el sistema a partir de una escala de tipo Likert, donde los usuarios califican su experiencia con el sistema (donde: 1 = muy insatisfecho, 5 = muy satisfecho).

Para implementar esta escala se hizo una breve encuesta de 6 preguntas que se podían calificar de 1 a 5

	Pregunta
Pregunta 1	Las imágenes que se me presentaron fueron relevantes para mis intereses, teniendo claro el contexto del conflicto armado
Pregunta 2	Percibí que el sistema mostró imágenes adaptadas al contexto y me permitió dar retroalimentación sobre mis preferencias
Pregunta 3	La relevancia en el contenido de las imágenes fueron adecuados para el contexto
Pregunta 4	Noté variedad en las imágenes mostradas, evitando repeticiones o contenido no relacionado con la temática.
Pregunta 5	Mi experiencia general al interactuar con el sistema fue satisfactoria
Pregunta 6	Al volver a hacer el recorrido el sistema mostró imágenes adaptadas a mis preferencias

Tabla 3: Tabla de preguntas al usuario. Fuente: Autor

Para esta encuesta se tuvo en cuenta si para la persona era su primer recorrido o si ya había hecho un segundo recorrido para lo cual solo un 30% hicieron un recorrido por segunda vez.

Cabe resaltar que las preguntas fueron formuladas enfocándose en un aspecto que se quería medir ya sea la relevancia para los usuarios, la personalización del contenido, que tan adecuado fue el contenido, que tan diverso fue el contenido y que tanto mejoró la experiencia al volver a usar la aplicación conforme a las imágenes.

Pregunta	Aspecto	Porcentaje
Las imágenes que se me presentaron fueron relevantes para mis intereses, teniendo claro el contexto del conflicto armado	Relevancia	74%
Percibí que el sistema mostró imágenes adaptadas al contexto y me permitió dar retroalimentación sobre mis preferencias	Personalización	86%
La relevancia en el contenido de las imágenes fueron adecuados para el contexto	Adecuación	76%

Noté variedad en las imágenes mostradas, evitando repeticiones o contenido no relacionado con la temática.	Diversidad	78%
Mi experiencia general al interactuar con el sistema fue satisfactoria	Experiencia	78%
Al volver a hacer el recorrido el sistema mostró imágenes adaptadas a mis preferencias	Adaptabilidad	87%

Tabla 4: Tabla de resultados por pregunta y aspecto a valorar. Fuente: Autor

7.2 Comparativa con el modelo MVMED+

A manera de comparación se tomó como referencia el articulo (Sun, Liu, & Mao, Multi-view learning for visual violence recognition with maximum entropy discrimination and deep features discrimination and deep features, 2018) el cual generó el modelo MVMED+ para el cual en dicho artículo se mostraron gráficas como:

Table 6
More prediction performance comparison between models with three views.

Method	Life			Scene			Physical		
	Recall	Precision	F1	Recall	Precision	F1	Recall	Precision	F1
CNN	61.95	73.84	65.90	85.93	82.90	83.65	73.09	89.65	79.14
SVM	77.92	76.99	77.45	94.15	94.45	94.29	91.17	91.30	91.17
KNN	55.95	**88.20**	68.44	77.89	**98.34**	86.92	71.37	95.29	71.37
LR	79.52	82.29	80.88	96.21	94.68	95.44	93.04	92.04	92.53
NB	75.15	83.24	78.98	92.30	95.19	93.72	85.94	**95.36**	90.40
MVMED+	**82.68**	86.38	**84.48**	**98.23**	95.12	**96.64**	**95.71**	93.72	**94.71**

Tabla 5: Tabla de comparación de los modelos presentada en el artículo de comparación. Fuente: "Multi-view learning for visual violence recognition with maximum entropy discrimination and deep features discrimination and deep features" de (Sun, Liu, & Mao, 2018)

Aquí se muestran las pedidas como en F1-SCORE y Precisión comparando el modelo MVMED+ contra otros modelos, por tal razón este se tomó como referencia ya que se manejaron estas dos mismas mediciones para el trabajo actual.

Con base en esto y debido a que no se logró obtener el mismo dataset que se trabajó en el articulo ya que no está disponible en el enlace mencionado en el artículo, se dispuso a hacer una validación del dataset actual (conflicto armado en Colombia) bajo este modelo con la finalidad de que el mismo dataset se trabajara en los dos modelos.

Para ello se implementó un código para la ejecución del modelo MVMED+ que tomara las imágenes que se trabajaron en este trabajo y se realizara la respectiva evaluación.

```python
# ----------------------------
# Dependencias opcionales (OpenCV/SIFT)
# ----------------------------
try:
    import cv2  # opcional
    _HAS_CV2 = True
    _HAS_SIFT = hasattr(cv2, "SIFT_create")
except Exception:
    _HAS_CV2 = False
    _HAS_SIFT = False


# ----------------------------
# Utilidades de imagen y vistas
# ----------------------------
def _leer_imagen(path, size=200):
    try:
        img = io.imread(path)
    except Exception as e:
        # Fallback a Pillow si imageio/skimage no puede abrir el archivo
        try:
            from PIL import Image
            with Image.open(path) as im:
                # convertir a RGB o L según el modo
                if im.mode in ("RGBA", "RGB"):
                    im = im.convert("RGB")
                    img = np.array(im).astype(np.float32) / 255.0
                else:
                    im = im.convert("L")
                    img = np.array(im).astype(np.float32) / 255.0
        except Exception:
            raise
    # Normalizar canales: aceptar RGB, RGBA o escala de grises
    if img.ndim == 3:
        # Si tiene 4 canales (RGBA), convertir a RGB primero
        if img.shape[2] == 4:
            try:
                img = color.rgba2rgb(img)
            except Exception:
                # fallback: descartar canal alfa
```

Ilustración 45; Código MVMED+ implementado para la comparación. Fuente: Autor

Como resultado de este ejercicio se obtuvo un 0.65 de pertinencia en el modelo lo cual indica que al evaluar el set de pruebas del conflicto armado en Colombia bajo este modelo se obtiene un valor inferior al que se estableció en el set de pruebas del articulo el cual era del 0.86 para la precisión,

```
[Test] Extrayendo vistas: 100% (1725/1725)
Métricas de evaluación:
  Accuracy: 0.6532
  Precision (weighted): 0.6532
  F1-score (weighted): 0.6496
```

Ilustración 46: Resultado de la ejecución del modelo MVMED+ con el dataset del conflicto armado. Fuente: Autor

En la tabla 6 se presentan los resultados obtenidos en una tabla comparativa entre el modelo MVMED+ versus el modelo planteado de MVL.

En esta tabla se presentan las dos métricas en común (Precisión y F1-Score), donde:

- MVMED+ Resultados obtenidos en el artículo: Muestra los datos obtenidos en el artículo de referencia como se muestra en la Tabla 5.
- MVMED+ Resultados obtenidos con DATASET DE CONFLICTO ARMADO: Muestra los resultados que al implementar el modelo MVMED+ genera con la data que se utilizó para el caso de estudio trabajado en este proyecto.
- Multiview learning propuesto - Resultados obtenidos con DATASET DE CONFLICTO ARMADO: Muestra los resultados del modelo MVL con la data que se utilizó para el caso de estudio trabajado en este proyecto.
- Diferencia entre modelos: Muestra la diferencia que hay entre los dos modelos con el mismo dataset.
- Porcentaje de mejora: Muestra un valor porcentual que representa la variación del diferencia entre un modelo y el otro.

Métricas	MVMED+ Resultados obtenidos en el artículo	MVMED+ Resultados obtenidos con DATASET DE CONFLICTO ARMADO	Multiview learning propuesto - Resultados obtenidos con DATASET DE CONFLICTO ARMADO	Diferencia entre modelos	Porcentaje de mejora
Precisión	86,38	65,32	82,28	16,96	13%
F1-score	84,48	64,96	76,62	11,66	12%

Tabla 6: Tabla de comparación entre resultados. Fuente: Autor

Por lo anterior, resulta necesario analizar las causas de la diferencia observada en los resultados. Una de ellas es el uso de un modelo de análisis de texto, el cual tiene un impacto significativo en la clasificación realizada en este proyecto, ya que su incorporación contribuye al incremento en el número de aciertos reflejados en los resultados.

Otra razón se relaciona con la arquitectura del modelo propuesto, que incorpora una capa de detección de objetos orientada a identificar elementos específicos como armas o personajes relevantes en el contexto, entre ellos líderes de grupos armados al margen de la ley, cuya presencia es altamente representativa. La detección de estos objetos permite asociarlos de manera directa con el conflicto armado, lo que fortalece la capacidad del modelo para discriminar imágenes de este tipo. En contraste, el modelo MVMED+ no

incorpora este tipo de información contextual, pues se limita al análisis de las características intrínsecas de las imágenes mediante su función de maximización. Si bien este enfoque es eficiente en un nivel general, su desempeño tiende a disminuir cuando se enfrenta a escenarios más específicos como el abordado en este caso de estudio.

7.3 Análisis de resultados

En este apartado se condensan los resultados obtenidos por las métricas, como se evidenció se manejaron 3 perspectivas que tienen diferentes enfoques.

La primera métrica evaluada, la Precisión, refleja el grado de acierto del sistema al momento de realizar la clasificación. En este caso, se obtuvo un valor del 82%, lo cual es considerado aceptable, ya que implica que, de cada diez imágenes procesadas bajo este modelo, aproximadamente ocho fueron clasificadas correctamente dentro del contexto del estudio. Si bien la expectativa para esta métrica era alcanzar un 85% o más, la diferencia no resulta significativa; sin embargo, sí evidencia la necesidad de continuar optimizando el modelo para mejorar su desempeño.

Con respecto a la segunda métrica (F1-score) la cual se enfoca en la precisión, pero sobre todo en la pertinencia de la clasificación al tener el parámetro de Recall, en esta métrica se obtuvo un 77% de precisión lo cual nos muestra que la sensibilidad que se maneja con el Recall hace que la precisión baje, siendo este un aspecto que se debe mejorar para lograr alcanzar la expectativa del 82% para esta métrica.

Y por último el análisis de la satisfacción del usuario para el cual logramos evidenciar un alto porcentaje en dos aspectos: el primero fue el aspecto de la Adaptabilidad en donde las personas que volvieron a hacer el recorrido tenían una muestra de imágenes basadas en la calificación anterior con un 87%, y el segundo en el aspecto de personalización en donde la pregunta va enfocada a cuan pertinente es la imagen al contexto (en este caso la temática del conflicto armado en Colombia) con un 86%. En general se obtuvieron resultados por encima del 75 % en todos los aspectos, pero esto da pie a seguir mejorando el sistema de clasificación, los conjuntos de datos y la forma de recolección de resultados.

Capítulo 8

Conclusiones

«Termina los proyectos que empieces, cumple los compromisos que hayas adquirido, has honor a tus promesas; si así lo haces, tanto tu yo consciente como el inconsciente podrán tener éxito, lo que llevará a un sentimiento de plenitud, valía e identidad»

John Roger.

8 Conclusiones

El trabajo realizado permite comprobar que el uso de un modelo basado en Multiview Learning puede representar una mejora significativa en la clasificación de imágenes en especial en contextos donde es necesario conocer un poco más del contexto sobre el cual se debe desenvolver el sistema de clasificación, tanto en términos de precisión técnica como de pertinencia del contenido presentado. La incorporación de múltiples vistas por imagen tanto visuales como semánticas enriqueció el análisis y permitió una normalización en la clasificación de los datos y por ende en la interpretación de cómo cada vista entiende las imágenes.

La aplicación de técnicas de aprendizaje supervisado, junto con el uso de métricas como precisión, recall y F1-score, permitió evaluar de forma objetiva el desempeño del modelo, que mostró resultados que por cada métrica están cerca de la expectativa que se tenía al iniciar el proceso de investigación. Además, la integración con un sistema de visualización confirmó la utilidad del modelo no solo presentar imágenes que están previamente clasificadas por este sistema de clasificación basado en Multiview learning, sino también para generar experiencias visuales más coherentes y relevantes desde el punto de vista del usuario.

8.1 Conclusiones

Los resultados obtenidos permitieron dar respuesta a las preguntas planteadas inicialmente

- Se lograron identificar elementos visuales clave como patrones visuales, objetos particulares y semántica específica para el contexto las cuales fueron tratadas como vistas individuales dentro de un modelo de clasificación basado en *Multiview Learning* permitiendo integrar la complejidad de las imágenes que se tomaron para el caso de estudio, esto a partir de la aplicación de los procesos de ingeniería de software aplicados desde la metodología.
- La combinación con técnicas de procesamiento de imágenes mediante el manejo en forma de capas independientes se logró mediante la estandarización de los

resultados que aportaba cada una de las vistas, permitiendo que se implementaran vistas complejas como redes neuronales convolucionales y detección de objetos lo cual enriqueció la capacidad de análisis del sistema y a su se tradujo en una mejora notable en la clasificación.

- Se incorporó la retroalimentación directa de los usuarios a partir de la integración con un sistema de visualización lo cual permitió que los usuarios brindaran una calificación sobre la imagen y estas valoraciones se pudieran implementar en el modelo como otra para tener una perspectiva más real y que mejoren la experiencia al entregar imágenes que cuenten con una validación de otros usuarios.
- Se establecieron los criterios de valoración del sistema de clasificación a nivel de métricas cuantitativas y de valoraciones de aspectos como la percepción de coherencia, la facilidad de navegación o la conexión contextual los cuales son claves para valorar su impacto real en entornos de visualización de contenido.
- Se logró establecer un punto de comparación entre una técnica de procesamiento de imágenes de Multiview Learning como lo es el MVMED+ aplicándolo sobre el mismo dataset que se planteó en el proyecto, esto permitió hacer esta comparativa de los resultados de los modelos evidenciando una merjor clasificación del modelo planteado de Multiview Learning frente el Modelo MVMED+.

8.2 Aporte al conocimiento

Más allá de sus resultados técnicos, este proyecto aporta una nueva perspectiva sobre el uso de Multiview learning (MVL) aplicado a imágenes que pueden tener un contexto especifico, un campo donde las interpretaciones no son siempre evidentes ni homogéneas. El modelo desarrollado no solo demostró ser eficaz, sino que también abrió la posibilidad de diseñar sistemas más sensibles al contexto y a la riqueza simbólica del contenido visual. Esta propuesta puede servir como base para futuras investigaciones en áreas donde el análisis visual requiere ir más allá de lo estructural y considerar también lo semántico, lo histórico y lo cultural.

8.3 Recomendaciones para futuros trabajos

Considerando los resultados obtenidos, como autor del trabajo se recomienda

1. Seguir incorporando nuevas técnicas de análisis visuales acordes a las tendencias (como lo puede ser inclusión de llm's) y hacer las variaciones en los casos de estudio para el mejoramiento de las clasificaciones.
2. Explorar nuevas formas de incorporar la interacción humana como parte activa del modelo, ya sea a través de retroalimentación explícita o mediante análisis del comportamiento del usuario.
3. Expandir el modelo con funcionalidades de integración otras fuentes de información, como textos descriptivos o referencias históricas, que funcionen como vistas adicionales para mejorar la comprensión del contenido.
4. Realizar en futuros trabajos podrían enfocarse en evaluar la experiencia del usuario de forma más integral, combinando datos cuantitativos con instrumentos cualitativos que permitan entender cómo se percibe y se valora realmente este tipo de clasificación en entornos no tan controlados.
5. Realizar un trabajo de investigación enfocado al planteamiento de una metodología para la curación de datasets de imágenes.

Capítulo 9

Bibliografía

Aquí recopilo brevemente los trabajos, los anhelos, los sueños, las vidas, los universos de muchas personas.
Que para mí fueron un referente, pero para la humanidad quedaran en lo eterno.

Fuente: Autor

9 Bibliografía

(s.f.).

Alegre, E., Pajares, G., & De la Escalera, A. (2016). *Conceptos y Métodos en Visión por Computador.* España. Recuperado el 19 de 04 de 2021, de https://intranet.ceautomatica.es/sites/default/files/upload/8/files/ConceptosyMetodosenVxC.pdf

Appflyer. (2018). *2018 App Retention Benchmarks [Data Study]*. Recuperado el 05 de 06 de 2019, de Appflyer: https://www.appsflyer.com/resources/2018-retention-benchmarks/

Atiyah1, B., & Hassan2, M. (06 de 06 de 2024). Machine learning and applications: A review study. *INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH AND ANALYSIS*, 2428-2432. Recuperado el 2025 de 04 de 11, de https://www.researchgate.net/publication/381166318_Machine_Learning_and_Applications_a_Review_Study

Audebert, N., Herold, C., Slimani, K., & Vidal, C. (15 de 07 de 2019). Multimodal deep networks for text and image-based document classification. *Computer Vision and Pattern Recognition*. Recuperado el 17 de 02 de 2022, de https://arxiv.org/abs/1907.06370

Borja-Robalino, R., Monleón-Getino, A., & Rodellar, J. (06 de 2020). Estandarización de métricas de rendimiento para clasificadores Machine y Deep Learning. *Iberian Journal of Information Systems and Technologies*, 184-196. Recuperado el 30 de 08 de 2024, de https://www.proquest.com/openview/a27e6b8c9b3aaa109da3a41f888ba8bc/1?pq-origsite=gscholar&cbl=1006393

Braakman, S., Pathmanthan, P., & Moore, H. (2022). Evaluation framework for systems models. CPT: Pharmacometrics & Systems Pharmacology. *CPT Pharmacometrics Syst*

Pharmacol, 264-289. Recuperado el 30 de 08 de 2024, de https://pubmed.ncbi.nlm.nih.gov/34921743/

Castaño Giraldo, E., Avella Estrada, A., Arango Murcia, A., & Sánchez, C. (2016). La imagen en el contexto de la violencia en Colombia: un acercamiento a distintas perspectiva. *Cuadernos de Música, Artes Visuales y Artes Escénicas*.

Castillo Aráuz, D., & Martínez, J. (15 de 12 de 2023). Predicción del rendimiento académico en la UNADECA por medio de sistemas de clasificación. *UNACIENCIA*, 17-35. Recuperado el 2024 de 09 de 14, de https://revistas.unac.edu.co/ojs/index.php/unaciencia/article/view/738

Centro de memoria historica de Colombia. (21 de 08 de 2025). *¡Basta ya! - Colombia memorias de guerra y dignidad*. Obtenido de Centro de memoria historica de Colombia: https://www.centrodememoriahistorica.gov.co/micrositios/informeGeneral

Diaz Monje, K. C. (2022). *Estudio y estructuración bibliográfica del estado del arte del uso de multiview learning en la clasificación de imágenes.* Bogotá. Recuperado el 30 de 04 de 2022

Du, C., Du, C., Huang, L., & He, H. (12 de 12 de 2018). Reconstructing Perceived Images From Human Brain Activities With Bayesian Deep Multiview Learning. *IEEE Transactions on Neural Networks and Learning Systems*, 1-14. Recuperado el 16 de 12 de 2021, de https://pubmed.ncbi.nlm.nih.gov/30561354/

Eco, U. (1976). *A Theory of Semiotics.* Bloomington: Indiana University Press. Recuperado el 20 de 08 de 2025, de https://raggeduniversity.co.uk/wp-content/uploads/2025/01/A-Theory-of-Semiotics-Umberto-Eco-1979.pdf

Fabbrizzi, S., Papadopoulos, S., Ntoutsi, E., & Kompatsiaris, I. (10 de 2022). A survey on bias in visual datasets. *Computer Vision and Image Understanding*. Recuperado el 21 de 08 de 2025, de https://www.sciencedirect.com/science/article/abs/pii/S1077314222001308

Farhani, G., Sica, R. J., & Daley, M. J. (17 de 02 de 2020). Classification of lidar measurements using supervised and unsupervised machine learning methods. *Atmospheric Measurement Techniques*, 391-402. Recuperado el 2024 de 09 de 14, de https://amt.copernicus.org/preprints/amt-2019-495/amt-2019-495.pdf

Fernández Marcos, S. (2023). *Sistema de reconocimiento de patrones en imágenes satelitales para la detección de objetos.* Universidad de Salamanca. Recuperado el 2024 de 02 de 03, de https://gredos.usal.es/bitstream/handle/10366/158423/Memoria.pdf?sequence=1&isAllowed=y

Francés Falip, J. (2020). *Redes neuronales convolucionales aplicadas a la identificación y medición automatizadas.* Barcelona. Recuperado el 25 de 06 de 2022, de https://upcommons.upc.edu/bitstream/2117/334221/2/tfg-javier-frances.pdf

gamco. (s.f.). *¿Qué es Red neuronal feedfordward?* Recuperado el 19 de 07 de 2024, de Gamco.es: https://gamco.es/glosario/red-neuronal-feedforward/

García, J., Gardel, A., Bravo, I., & Lázaro, J. L. (03 de 08 de 2014). Multiple View Oriented Matching Algorithm for People Reidentification. *IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS*, 1841 - 1851. Recuperado el 20 de 11 de 2020, de https://ieeexplore.ieee.org/abstract/document/6834787

Gelvez, J. (2025). Predicting police and military violence: evidence from Colombia and Mexico using machine learning models. *Humanities and social sciences communications*, 765. Recuperado el 21 de 08 de 2025, de https://www.nature.com/articles/s41599-025-04967-w

Gutiérrez Sanín, F. (2019). *Clientelistic Warfare: Paramilitaries and the State in Colombia (1982–2007) (Sociología Política Para los Desafíos del Siglo Xxi).* Recuperado el 21 de 08 de 2025, de https://www.buscalibre.com.co/libro-clientelistic-warfare-paramilitaries-and-the-state-in-colombia-1982-2007-sociologia-politica-para-los-desafios-del-siglo-xxi/9781787073654/p/51248941

Huang, A., Wang, Z., Zheng, Y., Zhao, T., & Lin, C.-W. (2021). Embedding Regularizer Learning for Multi-View Semi-Supervised Classification. 6997-7011. Recuperado el 15 de 01 de 2022, de https://pubmed.ncbi.nlm.nih.gov/34357859/

Jia, K., Lin, J., Tan, M., & Tao, D. (25 de 04 de 2019). Deep Multi-View Learning using Neuron-Wise Correlation-Maximizing Regularizers. Recuperado el 29 de 01 de 2021, de https://arxiv.org/abs/1904.11151

KrianVoleti. (27 de 01 de 2024). *Aprendizaje profundo: el algoritmo de optimización Adam para marketing*. Recuperado el 19 de 07 de 2024, de KrianVoleti.com: https://kiranvoleti.com/deep-learning-the-adam-optimization-algorithm-for-marketing

Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). ImageNet Classification with Deep Convolutional. *Neural Networks, 25*, 9. Recuperado el 15 de 07 de 2024, de https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf

Largo Garcia, C., & Marin Mazo, E. (03 de 2009). *Guía técnica para la evaluación de software.* Recuperado el 20 de 10 de 2024, de https://jrvargas.wordpress.com/wp-content/uploads/2009/03/guia_tecnica_para_evaluacion_de_software.pdf

Li, A., Chen, J., Chen, D., & Sun, G. (2020). Multiview Similarity Learning for Robust Visual Clustering. *Computer Vision – ACCV 2020 Workshops: 15th Asian Conference on Computer Vision, Kyoto, Japan*, (págs. 168-183). Tokio. Recuperado el 19 de 08 de 2021, de https://www.tib.eu/en/search/id/springer:doi~10.1007%252F978-3-030-69756-3_12/Multiview-Similarity-Learning-for-Robust-Visual?cHash=e4adf1d7f35ad82b0a1163de957b9e7a

Liu, B., Yu, A., Yu, X., Wang, R., Gao, K., & Guo, W. (09 de 09 de 2021). Deep Multiview Learning for Hyperspectral Image Classification. *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*, 7758 - 7772. Recuperado el 16 de 12 de 2021, de https://ieeexplore.ieee.org/document/9254125

Liu, M., Luo, Y., Tao, D., Xu, C., & Wen, Y. (2015). Low-Rank Multi-View Learning in Matrix Completion for Multi-Label Image Classification. *Twenty-Ninth AAAI Conference on Artificial Intelligence, 29*(1). Recuperado el 15 de 01 de 2021, de https://ojs.aaai.org/index.php/AAAI/article/view/9547

Liu, Y., Sun, S., & Yin, M. (2018). Multi-view Learning and Deep Learning for Microscopic Neuroblastoma Pathology Image Diagnosis. *Pacific Rim International Conference on Artificial Intelligence*, (págs. 545-558). Recuperado el 11 de 06 de 2021, de https://link.springer.com/chapter/10.1007/978-3-319-97304-3_42

Mann, S., Furness, T., Yuan, Y., Iorio, J., & Wang, Z. (2018). All Reality: Virtual, Augmented, Mixed (X), Mediated (X,Y), and Multimediated Reality. 14. Recuperado el 10 de 05 de 2020, de https://arxiv.org/pdf/1804.08386.pdf

Métricas de evaluación. (21 de 11 de 2024). Recuperado el 28 de 03 de 2025, de https://learn.microsoft.com/: https://learn.microsoft.com/es-es/azure/ai-services/language-service/custom-text-classification/concepts/evaluation-metrics

Mohamed, A. H., Hemeida, A. M., & Hassan, M. R. (2022). *Image classification based deep learning: A Review.* Aswan University Journal of Sciences and Technology. Recuperado el 01 de 09 de 2024, de https://www.researchgate.net/publication/366904831_Image_classification_based_deep_learning_A_Review

Mosleh, M., Al-Khulaidi, N., Gumaei, A. H., Alsabry, A., & Musleh, A. (07 de 08 de 2024). Classification and Evaluation Framework of Automated testing tools for agile software: Technical Review. *2024 4th International Conference on Emerging Smart Technologies and Applications (eSmarTA)*. Recuperado el 18 de 03 de 2025, de https://ieeexplore.ieee.org/document/10638902

Oliver, J. N. (2019). *Multi-View Imaging: Depth Estimation And Enhancement".* Recuperado el 27 de 09 de 2021, de https://dspace.uib.es/xmlui/handle/11201/150580

O'Shea, K., & Nash, R. (02 de 12 de 2015). An Introduction to Convolutional Neural Networks. *Neural and Evolutionary Computing*. Recuperado el 03 de 12 de 2022, de https://arxiv.org/pdf/1511.08458

Paula, L., Eckert, K., & Guismín, G. (09 de 2018). Clasificación de hojas de té al ingreso del proceso de secado mediante redes neuronales con datos supervisados y no supervisados. *X Congreso de AgroInformática (CAI 2018)*, 76-89. Recuperado el 19 de 07 de 2023, de http://sedici.unlp.edu.ar/handle/10915/70998

Pécaut, D. (2001). *Guerra contra la sociedad.* Espasa hoy. Obtenido de http://www.mamacoca.org/docs_de_base/La_Representacion_Social_del_narcotrafico/PecautDaniel_Guerra_Contra_la_Sociedad_EdEspasa_2001.pdf

Rettberg, A., & Prieto, J. (2018). Conflicto crudo: Petróleo, conflicto armado y criminalidad en Colombia. Recuperado el 21 de 08 de 2025, de https://www.researchgate.net/publication/326217676_Conflicto_crudo_Petroleo_conflicto_armado_y_criminalidad_en_Colombia

Romero, M. (2007). *Parapolítica. La ruta de la expansión paramilitar y los acuerdos políticos.* Recuperado el 21 de 08 de 2025, de https://bibliotecadigital.udea.edu.co/server/api/core/bitstreams/8bd59414-58a5-4e5d-8acb-9276d9e5c378/content

Sánchez Álvarez, R. (28 de 05 de 2021). Clasificación no supervisada de imágenes médicas y minería de datos. Algoritmo S3 vs K-medias. *Revista cubana de Investigaciones Biomédicas*, 40. Recuperado el 2023 de 09 de 14, de https://revibiomedica.sld.cu/index.php/ibi/article/view/1614/1008

Sánchez Cruz, A. F. (2025). *DISEÑO DE UN METAMODELO DE SOFTWARE PARA EL DESARROLLO DE APLICACIONES INMERSIVAS.* Bogotá. Recuperado el 2025 de 05 de 01

Sánchez Gómez, G., & Perea Restrepo, C. (2006). Guerras, Memoria e Historia. (L. C. Editores, Ed.) Obtenido de

http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0121-47052006000300009#(1)

Sánchez, G. (2007). Los estudios sobre la violencia: balance y perspectivas. *Pasado y presente de la violencia en Colombia*, 17-32.

Seeland, M., & Mäder, P. (26 de 12 de 2020). Multi-view classification with convolutional neural networks. *PLOS ONE*. Recuperado el 06 de 04 de 2021, de https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0245230

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*. Recuperado el 08 de 10 de 2023, de https://www.sciencedirect.com/science/article/abs/pii/S0306457309000259

Spencer, B., Hoskere, V., & Narazaki, Y. (04 de 2019). Advances in Computer Vision-Based Civil Infrastructure Inspection and Monitoring. *Engineering, 5*, 199-222. Recuperado el 2025 de 09 de 2, de https://www.sciencedirect.com/science/article/pii/S2095809918308130

Su, H., Maji, S., Kalogerakis, E., & Miller, E. L. (2015). Multi-view Convolutional Neural Networks for 3D Shape Recognition. *arXiv*. Recuperado el 24 de 11 de 2019, de https://arxiv.org/abs/1505.00880

SUN, L., WANG, J., HU, Z., XU, Y., & Cui, Z. (30 de 09 de 2019). Multi-View Convolutional Neural Networks for Mammographic Image Classification. Recuperado el 20 de 02 de 2021, de https://www.researchgate.net/publication/335595117_Multi-View_Convolutional_Neural_Networks_for_Mammographic_Image_Classification

Sun, S., & Chao, G. (08 de 2013). Multi-View Maximum Entropy Discrimination. Recuperado el 17 de 09 de 2020, de https://www.ijcai.org/Proceedings/13/Papers/253.pdf

Sun, S., Liu, Y., & Mao, L. (21 de 09 de 2018). Multi-view learning for visual violence recognition with maximum entropy discrimination and deep features discrimination

and deep features. *Information Fusion*, 43-53. Recuperado el 24 de 05 de 2019, de https://www.sciencedirect.com/science/article/abs/pii/S1566253518302616

Symflower. (s.f.). *Evaluating LLMs: complex scorers and evaluation frameworks*. Recuperado el 28 de 03 de 2025, de https://symflower.com/: https://symflower.com/en/company/blog/2024/llm-complex-scorers-evaluation-frameworks/

TensorFlow. (24 de 03 de 2022). *TensorFlow Core - Documentación esencial*. Obtenido de https://www.tensorflow.org/guide?hl=es-419

Tuozhong, Y., Wenfeng, W., & Yuhong, G. (12 de 11 de 2020). A Deep Multiview Active Learning for Large-Scale Image Classification. *Mathematical Problems in Engineering*. Recuperado el 30 de 01 de 2020, de https://www.hindawi.com/journals/mpe/2020/6639503/

Veit, A., Matera, T., Neumann, L., Matas, J., & Belongie, S. (19 de 01 de 2016). COCO-Text: Dataset and Benchmark for Text Detection and Recognition in Natural Images. *Computer Vision and Pattern Recognition*. Recuperado el 30 de 10 de 2023, de https://arxiv.org/abs/1601.07140

Ward Powers, D. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. *Journal of Machine Learning Technologies*, 2229-3981. Recuperado el 28 de 03 de 2025, de https://www.researchgate.net/publication/276412348_Evaluation_From_precision_recall_and_F-measure_to_ROC_informedness_markedness_correlation

Wong, K., Sinkovskaya, E., Abuhamad, A., & Mahmood, T. (23 de 03 de 2021). Multiview and Multiclass Image Segmentation using Deep Learning in Fetal Echocardiography. *Proceedings of SPIE - The International Society for Optical Engineering*. Recuperado el 15 de 11 de 2021, de https://www.tib.eu/en/search/id/spie:doi~10.1117%252F12.2582191/Multiview-

and-multiclass-image-segmentation-using?cHash=8f0b06bedd35889f260a0383c8224fb2

Wu, D., Zhang, C., Ji, L., Ran, R., Wu, H., & Xu, Y. (25 de 04 de 2021). Forest Fire Recognition Based on Feature Extraction from Multi-View Images. *Traitement du Signal*, 775-783. Recuperado el 20 de 11 de 2021, de https://www.researchgate.net/profile/Zhang-Chunjiong/publication/353490623_Forest_Fire_Recognition_Based_on_Feature_Extraction_from_Multi-View_Images/links/610919e6169a1a0103d4d1e4/Forest-Fire-Recognition-Based-on-Feature-Extraction-from-Multi-View-Images

Wu, F., Jing, X. Y., You, X., Yue, D., Hu, R., & Yang, J. Y. (21 de 08 de 2015). Multi-view low-rank dictionary learning for image classification. *Pattern Recognition*, 143-154. Recuperado el 27 de 05 de 2021, de https://www.sciencedirect.com/science/article/abs/pii/S0031320315003003

Xu, C., Dacheng, T., & Xu, C. (2013). A Survey on Multi-view Learning. *arXiv.org*. Recuperado el 18 de 07 de 2019, de https://arxiv.org/abs/1304.5634

Yan, C., Gong, B., Wei, Y., & Gao, Y. (01 de 02 de 2020). Deep Multi-View Enhancement Hashing for Image Retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Recuperado el 08 de 10 de 2021, de https://www.researchgate.net/publication/339480730_Deep_Multi-View_Enhancement_Hashing_for_Image_Retrieval

Yu, J., Li, J., Yu, Z., & Huang, Q. (03 de 12 de 2020). Multimodal Transformer with Multi-View Visual Representation for Image Captioning. *IEEE Transactions on Circuits and Systems for Video Technology*, 4467–4480. Recuperado el 17 de 11 de 2021, de https://dl.acm.org/doi/10.1109/TCSVT.2019.2947482

Zhang, C., Cheng, J., & Tian, Q. (16 de 08 de 2019). Multi-View Image Classification With Visual, Semantic and View Consistency. *IEEE TRANSACTIONS ON IMAGE*

PROCESSING, 29, 617 - 627. Recuperado el 21 de 12 de 2020, de https://ieeexplore.ieee.org/document/8804379

Zhang, L., Shum, H., Liu, L., Guo, G., & Shao, L. (2019). Multiview discriminative marginal metric learning for makeup face verification. *Neurocomputing*, 339-350. Obtenido de https://www.sciencedirect.com/science/article/pii/S0925231218314577

Zhang, X., Cheng, J., Xu, C., Lu, H., & Ma, S. (03 de 07 de 2009). Multi-view multi-label active learning for image classification. *2009 IEEE International Conference on Multimedia and Expo*. Recuperado el 07 de 03 de 2020, de https://ieeexplore.ieee.org/abstract/document/5202484

Zhao, J., Xijiong, X., Xin, X., & Shiliang, S. (2017). Multi-view learning overview: Recent progress and new challenges and new challenges. *Information Fusion*, 43-54. Recuperado el 09 de 04 de 2021, de https://www.sciencedirect.com/science/article/abs/pii/S1566253516302032

Zhao, Y., Cheung, Y.-m., You, X., Peng, Q., Peng, J., Yuan, P., & Shi, Y. (03 de 03 de 2021). Hyperspectral Image Classification via Spatial Window-Based Multiview Intact Feature Learning. *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*, 2294 - 2306. Recuperado el 03 de 12 de 2021

Zhu, S., Sun, X., & Jin, D. (08 de 02 de 2016). Multi-view semi-supervised learning for image classification. *Neurocomputing*, 136-142. Recuperado el 09 de 10 de 2020, de https://www-sciencedirect-com.bdigital.udistrital.edu.co/science/article/pii/S0925231216304416